SEARCH REQUEST FORM

Examiner # (Mandatory): _____   Requester's Full Name: _____

Art Unit _____ Location (Bldg/Room#): _____9E07_____ Phone (circle 305 306 308) _____

Serial Number: _____   Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

_____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms):

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

STAFF USE ONLY

Searcher: _____D. Schreiber_____

Searcher Phone #: _____308-4292_____

Searcher Location: _____CM1 2.E19_____

Date Picked Up: _____7/13_____

Date Completed : _____7/16_____

Clerical Prep Time: _____8_____

Terminal Time: _____11_____

Number of Databases: _____7_____

Type of Search		Vendors (include cost where applicable).	
_12_	N.A. Sequence	_____	STN
_1_	A.A. Sequence	_____	Questel/Orbit
_____	Structure (#)	_____	Lexis/Nexis
_____	Bibliographic	_____	WWW/Internet
_____	Litigation1	_√_	In-house sequence systems (list)
_____	Fulltext	_____	Dialog
_____	Procurement	_____	Dr. Link
_____	Other	_____	Westlaw
		_____	Other (specify)

MPsrch
Compugen

THIS PAGE BLANK (USPTO)

IMPORTANT INFORMATION ABOUT YOUR SEQUENCE SEARCH:

Compugen Sequence searching hardware and software explained:

This is the new sequence searching system that is currently being phased into as a replacement for the Maspar/Mpsrch platform. This system has been tested by both searchers and examiners, and has shown equivalent results to the Maspar system for the same databases. The results output format for all Compugen printed results are essentially the same **except for translations.**

Translation searching on Compugen explained:

The Compugen system utilizes Framesearch software for translations of proteins to nucleotides, and nucleotides to proteins. Some examiners have found these to be superior to the backtranslate software on Maspars.

FrameSearch searches a group of protein sequences for similarity to one or more nucleotide query sequences, or searches a group of nucleotide sequences for similarity to one or more protein query sequences. For each sequence comparison, the program finds an optimal alignment between the protein sequence and the corresponding codons on each the nucleotide sequence. Optimal alignments may include reading frame shifts. Please see any of the professional searching staff if you need assistance with this format.

File extensions for Compugen results transferred to floppy disks.

Compugen system search results will be delivered in one of two possible formats:
1. Standard concatenated files with .flp extension.
2. Compressed .zip files which decompressed yield two files as described below:

US08123456.**cmr** - Contains all commercial databases, may include Issued
US08123456.**pen** - Contains pending file results only

VERY IMPORTANT NOTE ABOUT PENDING FILE SEARCHES.

If your search contains file names with the following bolded extensions:
US08123456.**rap**   US08123456.**rnp**

Do not leave this search in the case, during prosecution, or after the case issues, since it contains pending data which is confidential.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 12:50:38 ; Search time 822.9 Seconds
                                           (without alignments)
                                           4334.446 Million cell updates/sec

Title: US-09-156-580-1
Perfect score: 997
Sequence: 1 cccagtgccatttttctct.........aaagttaattaaaaaaaa 997

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database :   GenEmbl:*
        1: gb_ba1:*
        2: gb_ba2:*
        3: gb_in:*
        4: gb_om:*
        5: gb_ov:*
        6: gb_pat:*
        7: gb_ph:*
        8: gb_pl1:*
        9: gb_pl2:*
       10: gb_prl:*
       11: gb_pr2:*
       12: gb_pr3:*
       13: gb_ro:*
       14: gb_st:*
       15: gb_sy:*
       16: gb_un:*
       17: gb_vi:*
       18: gb_htg:*
       19: em_ba:*
       20: em_fun:*
       21: em_hum1:*
       22: em_hum2:*
       23: em_in:*
       24: em_om:*
       25: em_ov:*
       26: em_or:*
       27: em_pat:*
       28: em_ph:*
       29: em_pl:*
       30: em_ro:*
       31: em_sy:*
       32: em_un:*
       33: em_vi:*
       34: em_htg:*
       35: em_sts:*
       36: gb_ba1:*
       37: gb_ba2:*
       38: gb_pl1:*
       39: gb_pl2:*
       40: gb_pr1:*
       41: gb_pr2:*
       42: gb_pr3:*
       43: gb_sts:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result NO.		Score	Query	Length	DB	ID	Description
1		89	8.9	1071	9	ATU38946	U38946 Arabidopsis
2		89	8.9	1071	39	ATU38946	U38946 Arabidopsis
3	c	76.4	7.6	99687	18	AC005956	AC005956 *** SEQUE
4	c	70.4	7.1	93845	9	ATAC004684	AC004684 Arabidops
5	c	70.4	7.1	93845	39	ATAC004684	AC004684 Arabidops
6		49	4.9	5931	2	DDU32174	U32174 Dictyosteli
7	c	47.2	4.7	38829	37	AE000787	AE000787 Borrelia
8	c	47.2	4.7	38829	9	AE000787	AE000787 Borrelia
9		46.8	4.7	1200	8	MISCVA	V00705 Yeast mitoc
10		46.8	4.7	1200	38	YSCMTVAR1	J01525 Yeast (S.ce
11		46.8	4.7	1200	38	MISCVA	J01525 Yeast (S.ce
12		46.2	4.6	1083	38	YSCMTVAR1	D26084 Petunia mRN
13		46.2	4.6	1083	38	PET2FDB2	D26084 Petunia mRN
14		46.2	4.6	1083	38	PET2FDB2	L77118 Methanococc
15		45.6	4.6	58407	37	MII2CG	L77118 Methanococc
16		45.6	4.6	58407	37	MII2CG	AE000664 Mus muscu
17		44.8	4.5	251364	13	MMAE000664	D26083 Petunia zin
18		44.6	4.5	3041	8	PET2FDB1	D26083 Petunia zin
19		44.6	4.5	3041	38	PET2FDB1	AF018638 Dictyoste
20		44.4	4.5	1453	3	DD1039PPR1	AF079445 Dictyoste
21		44.4	4.4	12668	3	AF079445	AL008972 Plasmodiu
22	c	44	4.4	28768	3	PFSC03020	AL010282 Plasmodiu
23		43.8	4.4	7467	3	PFSC04099	D13228 Yeast PHO8:
24		43.8	4.4	5100	38	YSCPHO81	D13228 Yeast PHO8:
25		43.8	4.4	5100	3	YSCPHO81	299773 Caenorhabd:
26	c	43.4	4.3	11442	3	CEH06A10	AF045453 Dictyoste
27		43.2	4.3	4374	3	AF045453	AF010544 Plasmodiu
28		42.2	4.2	553	3	AF010544	AF033931 Bactrocer
29		42.2	4.2	1139	3	AF033931	AL010153 Plasmodiu
30		42.2	4.2	6265	8	PFSC03080	M87278 Dictyosteli
31		42.2	4.2	3478	8	DDIADCYG	X66483 D.discoideu
32		42.2	4.2	3478	38	DDIADCYG	X53020 P.falciparu
33		41.8	4.2	3700	3	DDGP80G	Z36752 Caenorhabdi
34		41.8	4.2	235	3	PFPPF297	U20432 Dictyosteli
35	c	41.6	4.2	38383	3	CEF35H8	X95275 P.falciparu
36		41.4	4.2	6115	3	DDD20432	X75544 P.falciparu
37	c	41.4	4.2	15421	3	PFCOMPIRA	G37866 emaa2 Plasm
38	c	41.4	4.2	3075	3	XPFRNAPOL	L05498 Dictyosteli
39		41.4	4.2	165	43	G37866	AC004413 Homo sapi
40		41.2	4.1	2107	11	DDIADCYA03	L05498 Dictyosteli
41	c	41.2	4.1	114827	38	AC004413	AC004413 Homo sapi
42		41.2	4.1	2107	3	DDIADCYA03	G37927 c3m61 Plasm
43	c	41.2	4.1	114827	8	AC004413	M38703 Dictyosteli
44		41.2	4.1	221	43	G37927	
45		41	4.1	2696	8	DDIPRKNGPK	

ALIGNMENTS

RESULT   1
ATU38946

```
LOCUS       ATU38946     1071 bp    DNA             PLN       29-NOV-1995
DEFINITION  Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.
ACCESSION   U38946
NID         g1079668
KEYWORDS    
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
  TITLE     Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries
  JOURNAL   Nature 378 (6553), 199-203 (1995)
  MEDLINE   96069740
REFERENCE   2  (bases 1 to 1071)
  AUTHORS   Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech,
            Pasadena, CA 91125, USA
FEATURES             Location/Qualifiers
```

```
source          1..1071
                /organism="Arabidopsis thaliana"
                /strain="Landsberg erecta"
                /db_xref="taxon:3702"
                /chromosome="3"
5'UTR           join(1..68,175..202)
                /gene="sup"
gene            1..1071
                /gene="sup"
mRNA            join(1..68,175..1071)
                /gene="sup"
intron          69..174
                /gene="sup"
CDS             203..817
                /gene="sup"
                /standard_name="SUPERMAN"
                /note="Allele: Ler wild type;  C2H2 zinc finger protein"
                /codon_start=1
                /function="flower-specific gene controlling the boundary
                of the stamen and carpel whorls"
                /product="SUPERMAN"
                /db_xref="PID:g1079669"
                /translation="MERSNSIELRNSFYGRARTSPWSYGYDNCQQDHDYLLGFSWPP
                RSYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSTPSPPYPNPNYSYSTM
                ANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSSLLVEAGE
                ATRFTSKDACKILRNDEIISLELEIGLINESEQDLDELRLGFA"
3'UTR           818..1071
                /gene="sup"
BASE COUNT      326 a    226 c    183 g    336 t
ORIGIN

Query Match           8.9%;       Score 89;  DB 9;  Length 1071;
Best Local Similarity 68.9%;      Pred. No. 3.7e-08;
Matches 122;  Conservative   0;  Mismatches  55;  Indels    0;  Gaps    0;

Qy   313 agccattacgagataattatggaaatgaagatcattacttggtggactattccttgg 372
         ||||  |||||| ||| |||| | |||||  ||||||| ||||| |||||||
Db   269 AGCTATGGAGATTATGATAATTGCCAACAGGATCATGATTATCTTCTAGGGTTTTCATGG 328

Qy   373 cctccaagatcttatacatgtagctttgtaaaagggaattagatctgctcaagctctt 432
         || |||||||| |||||| |||||||  |||||||| |||||| ||||| ||||||
Db   329 CCACCAAGATCCTACACTTGCAGCTTCTGCAAAAGGGAATTCAGATCGGCTCAAGCACTT 388

Qy   433 ggtggacacatgaatgttcatagagagatagagccatttttgagacaatcaccacct 489
         ||||||  ||||| |||||| ||||||||| ||||||| | | |  |||| || ||
Db   389 GGTGGCCACATGAATGTTCACAGAGAGAGACAGCAGCAAGACTCAGATTACAACAGTCT 445


RESULT 2
ATU38946
LOCUS       ATU38946    1071 bp    DNA            PLN       29-NOV-1995
DEFINITION  Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.
ACCESSION   U38946
NID         g1079668
KEYWORDS    .
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
  TITLE     Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries
  JOURNAL   Nature 378 (6553), 199-203 (1995)
  MEDLINE   96069740
REFERENCE   2  (bases 1 to 1071)
  AUTHORS   Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech,
            Pasadena, CA 9125, USA
FEATURES             Location/Qualifiers
source          1..1071
```

```
                /organism="Arabidopsis thaliana"
                /strain="Landsberg erecta"
                /db_xref="taxon:3702"
                /chromosome="3"
5'UTR           join(1..68,175..202)
                /gene="sup"
gene            1..1071
                /gene="sup"
mRNA            join(1..68,175..1071)
                /gene="sup"
intron          69..174
                /gene="sup"
CDS             203..817
                /gene="sup"
                /standard_name="SUPERMAN"
                /note="Allele: Ler wild type;  C2H2 zinc finger protein"
                /codon_start=1
                /function="flower-specific gene controlling the boundary
                of the stamen and carpel whorls"
                /product="SUPERMAN"
                /db_xref="PID:g1079669"
                /translation="MERSNSIELRNSFYGRARTSPWSYGYDNCQQDHDYLLGFSWPP
                RSYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSTPSPPYPNPNYSYSTM
                ANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSSLLVEAGE
                ATRFTSKDACKILRNDEIISLELEIGLINESEQDLDELRLGFA"
3'UTR           818..1071
                /gene="sup"
BASE COUNT      326 a    226 c    183 g    336 t
ORIGIN

Query Match           8.9%;       Score 89;  DB 39;  Length 1071;
Best Local Similarity 68.9%;      Pred. No. 3.7e-08;
Matches 122;  Conservative   0;  Mismatches  55;  Indels    0;  Gaps    0;

Qy   313 agccattacgagataattatggaaatgaagatcattacttggtggactattccttgg 372
         ||||  |||||| ||| |||| | |||||  ||||||| ||||| |||||||
Db   269 AGCTATGGAGATTATGATAATTGCCAACAGGATCATGATTATCTTCTAGGGTTTTCATGG 328

Qy   373 cctccaagatcttatacatgtagctttgtaaaagggaattagatctgctcaagctctt 432
         || |||||||| |||||| |||||||  |||||||| |||||| ||||| ||||||
Db   329 CCACCAAGATCCTACACTTGCAGCTTCTGCAAAAGGGAATTCAGATCGGCTCAAGCACTT 388

Qy   433 ggtggacacatgaatgttcatagagagatagagccatttttgagacaatcaccacct 489
         ||||||  ||||| |||||| ||||||||| ||||||| | | |  |||| || ||
Db   389 GGTGGCCACATGAATGTTCACAGAGAGAGACAGCAGCAAGACTCAGATTACAACAGTCT 445


RESULT 3
AC005956/c
LOCUS       AC005956   99687 bp    DNA            HTG       04-DEC-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Arabidopsis thaliana 'MITSUI' P1
            'MHK10' genomic sequence near marker 'EG05C12'; HTGS phase 2, 1
            ordered pieces.
ACCESSION   AC005956
NID         g3962504
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 99687)
  AUTHORS   Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M., Shen,M.,
            Ronning,C.M., Rounsley,S.D., Fraser,C.M., Somerville,C.R. and
            Venter,J.C.
  TITLE     Arabidopsis thaliana 'MITSUI' BAC 'MHK10' genomic sequence near
            marker 'EG05C12'
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99687)
  AUTHORS   Rounsley,S.D. and Lin.X.
  TITLE     Direct Submission
```

JOURNAL   Submitted (10-NOV-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
COMMENT   On Dec 4, 1998 this sequence version replaced gi:3894154.
            ***
            *** WARNING: Phase 2 High Throughput Genome Sequence ***
            ***
            * This sequence is unfinished. It consists of 1 contigs for
            * which the order is known. The lengths of the gaps have been
            * estimated by the submitter but are not known exactly. When
            * sequencing is complete, the sequence data presented in this
            * record will be replaced by a single finished sequence
            * with the same accession number.
            *   1 99687: contig of 99687 bp in length.
FEATURES   Location/Qualifiers
     source   1..99687
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="MHK10"
BASE COUNT   30477 a 18393 c 19273 g 31544 t
ORIGIN

Query Match   7.6%;   Score 76.2;   DB 18;   Length 99687;
Best Local Similarity   72.3%;   Pred. No. 3.8e-06;
Matches   99; Conservative   0; Mismatches 38; Indels   0; Gaps   0;

Qy   355 ggtggactattctcttggcctccaagatcttatacatgtacttttgtaaaagggaattt 414
           ||| || |||| | |||||||| ||| ||||||||||||||| ||||||||||||||||
Db 45805 GCTGGGCTGATCCCATGGCCTCCCAAGAATTACACTTGCCAGCTTCTGCGAGGAGAGTTC 45746

Qy   415 agatccgtcaagctcttggtggaccacatgaatgttcatagaagagatagagccatttg 474
           |||||||||||||| |||||| |||||| |||||| |||||||||| |||| | |||||
Db 45745 AGATCTGCTCAAGCACTTGGAGGAGGCCACATGAATGTTCATAGAAGACACGAGCAAAACTC 45686

Qy   475 agacaatcaccacctag 491
           ||| || |||| | ||
Db 45685 AGGCAGATCCCTTCTTG 45669

RESULT   4
ATAC004684/c
LOCUS   ATAC004684 93845 bp   DNA   PLN   17-JUN-1998
DEFINITION Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence,
            complete sequence.
ACCESSION   AC004684
NID   g3236234
KEYWORDS   HTG.
SOURCE   thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 93845)
  AUTHORS   Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,
            Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE   Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 93845)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3 (bases 1 to 93845)
  AUTHORS   Rounsley,S.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT   On Jun 17, 1998 this sequence version replaced gi:3172168.
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr,

            Rockville, MD 20850,
            USA
            e-mail: rounsley@tigr.org
            BAC clone F13M22 is from Arabidopsis chromosome II and is near the
            molecular marker ve018.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge).
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES   Location/Qualifiers
     source   1..93845
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="ve018"
                     /clone="F13M22"
     misc_feature   complement(10..29)
     misc_feature   168..249
                     /note="exon predicted by xgrail, quality marginal"
     mRNA   <464..>619
                     /note="exon predicted by xgrail, quality good_shadowexon"
     gene   <464..>619
                     /gene="F13M22.1"
                     /note="predicted by genscan"
     CDS   <464..619
                     /gene="F13M22.1"
                     /gene="F13M22.1"
                     /note="hypothetical protein, 5' partial"
                     /codon_start=1
                     /db_xref="PID:g3236250"
                     /translation="SRGDVRVEQRRTHKDGEDKRLLDPATRRKLETNAITASGTKGF
                     EIGNGNL"
     mRNA   complement(join(835..1212,1373..1513,1600..1693,
                     1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
                     2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
                     4167..4353,4486..4549,4633..4842,4938..5006,5088..>5654))
                     /gene="F13M22.2"
     gene   complement(835..>5654)
                     /gene="F13M22.2"
     CDS   complement(join(1099..1212,1373..1513,1600..1693,
                     1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
                     2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
                     4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
                     /gene="F13M22.2"
                     /note="unknown protein"
                     /codon_start=1
                     /db_xref="PID:g3236235"
                     /translation="MPKEENGQLKRDRLDDTDEGNKGDHFPSKKQAKEASNDDITSE
                     ISNPVASPVSTSLFRDVSSQPVKSGLVECSGSDFGSEETVSDDASVVGSSQTEQSSD
                     VLPSRFVLEIPKHLSSTGITKITFKLSKPKKEFDDLPLIKDHTWDAGVVKMPKKIVS
                     LSYPSNVKKLLETGILEGARVKYISTPPVROLLGIIHSGGYLCGCTCNFSKVLSAYE
                     FEQHAGAKTHRPNNHIFLENRRAVYNIVQELKTAPRVVLEEVIRNVAGSALNEEGLRA
                     WKGDGQQVSNVIACLTGIILRIPLSVNALLSYHFVLLSETDSCESCSYLGPGLDES

QSLTPCSVENHYFSEKTYAKDTLDEPKRIAKKLTSHVSGTGCHKKVSEGSNRKRDNDL
HRLLFMPNGLPDGTELAYYVKTQISPSQFEAHAGMAARRQPYRHIFISSGLSLHDIAM
SLANGHVITTGDSDDMCSICGDGGDLLLCAGCPQAFHTACLKFQSMPEGTWYCSSCND
GPISSKKATTTDPSGNARPIVIRLSRVVKAPESDIGGCVFCRSHDFSIGKFDDRTVIL
CDQCEKEYHVGCLRENGFCDLKEIPQEKWFCCSNCSRIHTAVQNSVSCGPQTLPTPLL
DMICRKDREKGIFTDIGDTVEWRILSGKSRYPEHLPLLSRAAVIFRECFDPIVAKSGR
DLJPVMVYGRNISGQEFGGMYCLVLIVNSLVSAALLRIFGQEVAELPIVATSREYQG
RGYFQGLYACVENLLSSLNVENLVLPAAEEAESIWTKKFGFTKMSDQQLQEYQKEVQL
TIFKGTSMLEKKVPKATTGLSESTTLI"

repeat_region	5010..6135
	/rpt_family="(GGAA)n"
misc_feature	6023..6202
	/note="exon predicted by xgrail, quality
	marginal_shadowexon"
repeat_region	6179..6236
	/rpt_family="POLY_A"
mRNA	join(<6647..6702,7161..7304,7322..>7499)
	/gene="F13M22.3"
gene	<6647..>7499
	/note="predicted by genscan"
CDS	join(6647..6702,7161..7304,7322..7499)
	/gene="F13M22.3"
	/note="hypothetical protein"
	/codon_start=1
	/db_xref="PID:g3236236"
	/translation="MTTWKKVREENKSFQFLFKFAWRIPHFVYGATWTLFLTLTVAII
	SLAPEFAFVSAIFPSSSEVFSRRHIGSAGSSLGGLVFTGESFPEVEDGPRCAAGFCSN
	CCGFRRRRKDYGPLGGRRSPLI"
mRNA	complement(join(<7761..7931,8019..8065,8154..8238,
	8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
	/gene="F13M22.4"
gene	complement(<7761..>9886)
	/gene="F13M22.4"
CDS	complement(join(7761..7931,8019..8065,8154..8238,
	8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
	/gene="F13M22.4"
	/codon_start=1
	/product="putative ribitol dehydrogenase"
	/db_xref="PID:g3236237"
	/translation="MGIYGVMTGKKGKSGFGSASTAEDVTQAIDASHLTAIIGGTSG
	IGLEAARVLAMRGAHVIAARNPKAANESKEMLQMNPNARVDYLQIDVSSIKSVRSF
	VDQFLALNVPLNLINNAGVMFCPFKLIEDGIESQFATNHIGHFLLINLLLDKMKSTA
	RESGVOGRIVNLSIIAHTYYSEGIKFOGINDAGPKGYSERRAYGQSKLLHSNALSR
	RLQEEGVNITINSVHPGLVTTNLFRYSGFSMKVFRAMTFLFWKNIPQGAATTCYVALH
	PDLEGVTGKYFGDCNIVAPSKFATNNSLADKLMDFSVFLIDSISK"
repeat_region	8658..8684
	/rpt_family="AT_rich"
repeat_region	complement(9842..9915)
	/rpt_family="(GA)n"
misc_feature	10503..10670
	/note="exon predicted by xgrail, quality
	marginal_shadowexon"
repeat_region	10725..10746
	/rpt_family="AT_rich"
mRNA	complement(join(11416..11526,11608..12116,12551..12628,
	12737..13604))
	/gene="F13M22.5"
gene	complement(11416..13604)
	/gene="F13M22.5"
CDS	complement(join(11610..12116,12551..12628,12737..13522))
	/gene="F13M22.5"
	/codon_start=1
	/product="putative ARF1 GTPase activating protein"
	/db_xref="PID:g3236238"
	/translation="MAAARRLRFLQSQPENKVCVDCSOKNPQWASISYGIFMCLECSG
	KHRGLGVHISFVRSVTMDSWEIQIKKMDAGGNERLNFLAQYGISRKETDIISKYNSN
	AASYRDRIQALAEGRQWRDPPIVKESVGGLMKADYGISRETDIISKYNSN
	DSFRSTDMRRNQSAGDFRSSGRGAPAKSKSSEDIYSRSQLEASAANKESFFAKRMAE
	NESKPEGLPPSQQKVGFGSSPGPAPRSNQQSGGGDVFSVMSEGFGRLSLVAASAAN
	VVQTGTMEFTSKVKEGGLDQTVSETVNVVASKTEIGQRTWGIMKGVMAIASQKVEEF
	TKEEASTWNQQNKTEGNGYYONSGIGNKTANSSFGGSOSSSSGHNNSYRNSNSWDDWG
	EENNSKKEAAPKVSTSNDDDDGWAGWDDNDAKDDDFYYQPASDKKKSVGHNGKSDTAW

TGGGFL"
11697..11757
/rpt_family="(CAT)n"
13716..13952

repeat_region	
misc_feature	/note="exon predicted by xgrail, quality
	excellent_shadowexon"
misc_feature	complement(14172..14309)
	/note="exon predicted by xgrail, quality marginal"
mRNA	join(14707..14911,15566..15765,15964,16057..16189,
	16311..16602,16755..17294)
	/gene="F13M22.6"
gene	14707..17294
	/note="identical to GB:U40269"
	/gene="F13M22.6"
CDS	join(15634..15765,15858..15964,16057..16189,16311..16602,
	16755..17182)
	/gene="F13M22.6"
	/codon_start=1
	/product="origin recognition complex protein"
	/db_xref="PID:g3236239"
	/translation="MEDIENIEEDEYGFSRNYFLAKELGGASKRSAHKLSDIHIVDEQ
	ELRETASTIEMKHSKEISELMSDYKTMYSKWVFELRCGFGLLMYGFGSKKALVEDFAS
	ASLTDYSVVVINGYLPSVNLKQVLALAELLSELLKCKRKSSGSLSKGQETFPSRSMD
	DILSFLHGPQSGDKDCFICVVVHNIDGPALRDPESOQTLARLSSCSHIRLVASIDHVN
	APLLWDKKMVHKQFNWLWHHVPTFAPINVEGVFFPLVLAQGSTAQTAKTAAIVLQSLT
	PNGQNVFKILAEYOLSHPDEDGMTDDLYSASRERFFVSSQVTLNSHLTEFKDHELVK
	TKRNSDGQECLNIPLTSDAIRQLLDLNQ"
mRNA	complement(<17532..>19016)
	/gene="F13M22.7"
gene	complement(<17532..>19016)
	/gene="F13M22.7"
CDS	complement(17532..19016)
	/gene="F13M22.7"
	/note="unknown protein"
	/codon_start=1
	/db_xref="PID:g3236240"
	/translation="MENHHPSTLLSMDSSASSHEEDLEMNNNRQSLLSGPPDINLPL
	SAERSPPPPWNLDACDILDVGLGSQAYTENYMSVVPKVGRKCAKRVDSIWGAWFFF
	SFYFKPALNEKSKAKIVRDSNGISGFDKSDLKLDVFLVQHDMENMYMWVFKERPENAL
	GKMQLRSYMNGHSRQGDRLFPFSVEKGFVRSHRMQRKHYRGLSNPQCVHGIELVPLPN
	LTCLDEEERKRWMELTGRDLNFTIPPEASDFGSWRNLPNTDFELERPTPSLKNPTANH
	SKKLLNGSGLNLSTQPSNHSNGEATDLSPSSHKKRKDLFSNGTHEEECCLTVNPQPPV
	IEAHQNELPTWSNEFTGAMKNVYGPVTAKTIYEDEEGLIIISLPFVDLNSVKVSWR
	NTLTHGIIYKVSCLSTSRVPFIKRHDRTFLTDSASEHCPGEFVREIPLSNRIPEDAN
	IEAYDGPGSVLEILVPKLRAGPEEHEVRVCLRPNLGGNDLMLT"
misc_feature	complement(19045..19091)
	/note="exon predicted by xgrail, quality marginal"
misc_feature	19455..19545
	/note="exon predicted by xgrail, quality
	excellent_shadowexon"
mRNA	join(<20811..21507,21667..21805,22028..22434,22931..23006,
	23118..23233,23331..>23788)
	/gene="F13M22.8"
gene	<20811..>23788
	/gene="F13M22.8"
CDS	join(20811..21507,21667..21805,22028..22434,22931..23006,
	23118..23233,23331..23788)
	/gene="F13M22.8"
	/codon_start=1
	/product="putative zinc finger protein"
	/db_xref="PID:g3236261"
	/translation="MFNNTTSFGSGPGIVVVPTPATTVPTTDFPGTTITSNSTFIII
	GPPPPFAPPRSIDLTPLKLIFVVIAFVAVPALVYALFFNGPCSSSRRNSSSSRTSSS
	SDDTPHATVDTPPITETTVTSESGGKFHKDTHSKEIGNECSVCLMVFTDSDELRQLSE
	CKHAFHVLCIETWKDHPNCPICRTDVSVTEAPNVPNVNRSGGNRRVSAT
	SRDDWRQGLPDASTTVSVSVRDVKSDVDSQAHPTVIEDLILIGRLLIHTRLPIRTL
	SFCSRKPSDSSSGLAPNRNLATKSTIPRFAYLVTGTKGDGKRVKRLLKAIHHPRNYYL
	LHLDLEASDEERMELAKYVRSEKKFENVMVMGLADLVTEKGPTMLASTLHGVAILLK
	KAKDWDWFINLSASDYPLMPQDDILHIFSYLPRYLNFIEHTSNIGWKENQRARPIIID
	PGFYHLKKSGVFWAKERRSLPASFKLFMGSTSVALTRPFLEFCIWGWDNLPRTLLMYY
	TNFLLSSEGYFQVTVVCNNKDYQNTTVNHDLHYTKWDPLQQRTLNVTVENFRDMVQSGA
	PFAREFREDDLVLDKIDIELLGQTDTGLELKTPDVVKPTVSWRLEKLMVRLLDHENF
	RAKQCK"

```
repeat_region   complement(20872..20971)
                /rpt_family="(TGG)n"
repeat_region   complement(21518..21543)
                /rpt_family="AT_rich"
misc_feature    complement(24200..24291)
                /note="exon predicted by xgrail, quality good shadowexon"
repeat_region   complement(25078..25159)
                /rpt_family="(TA)n"
mRNA            join(<25358..25408,25622..>26563)
                /gene="F13M22.9"
gene            <25358..>26563
                /gene="F13M22.9"
CDS             join(25358..25408,25622..26563)
                /gene="F13M22.9"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236241"
                /translation="MVFSSIQAYLDSSWQQAPPSNYNHDGTGASANGHVLRPQLQP
                QQPQQQPHPNGSGGGGGGGSIRAGSMVDRARQANVALPEAALKCPRCESTNTKFC
                YFNNYSLTQPRHFCKTCRRYWTRGGALRNVPVGGGCRRNRRTKSNNNNNSTATSNN
                TSFSSGNASTISTILSSHYGGNQESILSQILSPARLMNPTYNHLGDLTSNTKTDNNMS
                LLNYGGLSQDLRSIHMGASGGSLMSCVDEWRSASYHQOSSMGGGNLEDSSNPNPSANG
                FYSFESPRITSASISSALASQFSSVKVEDNPYKWVNVNGNCSSWNDLSAFGSSR"
repeat_region   25591..25728
                /rpt_family="(CAG)n"
repeat_region   26002..26054
                /rpt_family="(CAA)n"
misc_feature    complement(27038..27122)
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   28179..28208
                /rpt_family="AT_rich"
misc_feature    complement(28259..28328)
                /note="exon predicted by xgrail, quality
                marginal shadowexon"
repeat_region   complement(28897..28980)
                /rpt_family="(CAG)n"
repeat_region   29086..29133
                /rpt_family="(TA)n"
misc_feature    complement(29945..30023)
                /rpt_family="AT_rich"
                /note="exon predicted by xgrail, quality good"
mRNA            complement(join(<30476..30515,30659..30858,31168..>31269))
                /gene="F13M22.10"
gene            complement(<30476..>31269)
                /gene="F13M22.10"
CDS             complement(join(30476..30515,30659..30858,31168..>31269))
                /gene="F13M22.10"
                /codon_start=1
                /product="putative ribosomal protein L36"
                /db_xref="PID:g3236242"
                /translation="MTTPAVKTGLFVGLNKGHVVTRELAPRPNSRKGKTSKRTIFIR
                KLIREVAGMAPYEKRITELLKVGKDKRALKVAKRKLGTHKRAKRKREEMSSVLRKMRS
                LGGAAAAEKKM"
misc_feature    complement(31377..31401)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    complement(join(<32935..33310,33630..>33739))
                /note="exon predicted by genscan"
repeat_region   complement(32006..32036)
                /note="exon predicted by xgrail, quality marginal"
repeat_region   32044..32161
                /rpt_family="(GAAA)n"
repeat_region   complement(32282..32354)
                /rpt_family="(GAAAA)n"
mRNA            32653..32718
                /rpt_family="POLY_A"
gene            complement(join(<32935..33310,33630..>33739))
                /gene="F13M22.11"
CDS             complement(join(32935..33310,33630..33739))
                /gene="F13M22.11"
                /codon_start=1
                /db_xref="PID:g3236243"
                /translation="MEIGNVKADVNLVLNLREAERHRYFDYFIDEILWLSFLGMSET

                SLKEEEVTLGGVKETREEECSDDKLVLDICLKRKRLDEEEEVVEEENDGFKTPTRP
                ENRIPIVRECPPAPMKRSSEMFRGRTMYCRRRLSFLPEDDVNSFITDLQWRTTTMTIK
                K"
repeat_region   complement(33140..33196)
                /rpt_family="(GAA)n"
repeat_region   33357..33398
                /rpt_family="(GAA)n"
repeat_region   33469..33489
                /rpt_family="AT_rich"
mRNA            complement(34248..34296)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    35206..35258
                /note="exon predicted by xgrail, quality good"
misc_feature    complement(35273..35341)
                /rpt_family="(GGAA)n"
repeat_region   join(35801..35886,36023..36416,36517..37130,37242..37322)
                /gene="F13M22.12"
mRNA            35801..37322
                /note="identical to GB:U29480"
                /gene="F13M22.12"
gene            join(35827..35886,36023..36416,36517..37130,37242..37307)
                /gene="F13M22.12"
                /codon_start=1
CDS             /product="actin 3 protein"
                /db_xref="PID:g3236244"
                /translation="MADGEDIQPLVCDNGTGMVKAGFAGDDAPRAVFPSIVGRPRHTG
                VMVGMGQKDAYVGDEAQSKRGILTLKYPIEHGIVNNWDDMEKIWHHTFYNELRVAPEE
                HPILLTEAPLNPKANREKMTQIMFETFNAPAMYVAIQAVLSLYASGRTTGIVLDSGDG
                VSHTVPIYEGYALPHAILRLDLAGRDLTDALMKILTERGYSFTTAEREIVRDIKEKL
                CYIALDYEQELETAKTSSSVEKNYELPDGQVITGSERFRCPEVLYQPSMIGMENAGI
                HETTYNSIMCDVDIRKDLYGNIVLSGGTMFPGIADRMSKEITALAPSSMKIKVVAP
                PERKYSVWIGGSILASLSTFQQMWIAKAEYDESGPSIVHRKCF"
repeat_region   37700..37817
                /rpt_family="(CAAA)n"
mRNA            complement(<37864..>38967)
                /gene="F13M22.13"
gene            complement(<37864..>38967)
                /gene="F13M22.13"
CDS             complement(37864..38967)
                /gene="F13M22.13"
                /codon_start=1
                /product="myb-related protein"
                /db_xref="PID:g3236245"
                /translation="MKERQRWSGEEDALLRAYVRQFGPREWHLVSERMNKPLNRDAKS
                CLERWKNYLKPGIKKGSLTEEEQRLVIRLQEKHGNKWKKIAAEVPGRTAKRLGKWWEV
                FKEKQROREEKESNKRVEPIDESKYDRILESFAEKLVKERSNVVPAAAAATVVMANSN
                GGFLHSEQOVQPPNPVIPPWLATSNGNNVVARPPSVTLTLSPSTVAAAPQPPIPWL
                QQQOPERAENGPGGLVLGSMPSCGSGSSESVFLSELVECCRELEGHRAWADHKKEAA
                WRLRRLELQESEKTCRQREKMEEIEAKNMALREOQKNAMEKIEGEYREQVLGLRRDA
                EAKDQKLADQWTSRHIRLRTKFLEOQMGCRLDRP"
repeat_region   complement(39531..39561)
                /rpt_family="AT_rich"
repeat_region   complement(40737..40818)
                /rpt_family="(TAA)n"
repeat_region   complement(41089..41112)
                /rpt_family="POLY_A"
repeat_region   complement(41190..41213)
                /rpt_family="POLY_A"
misc_feature    complement(41283..41310)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   complement(41558..41602)
                /rpt_family="AT_rich"
repeat_region   complement(41966..42068)
                /rpt_family="(GAAA)n"
repeat_region   complement(42281..42325)
                /rpt_family="AT_rich"
misc_feature    42889..42968
                /note="exon predicted by xgrail, quality excellent"
misc_feature    43133..43154
                /note="exon predicted by xgrail, quality good"
repeat_region   43371..43479
```

```
mRNA            /rpt_family="(TAAAA)n"
                complement(join(43750. .44452,45240. .45558,45719. .45963))
                /gene="F13M22.14"

gene            complement(43750. .45963)
                /gene="F13M22.14"

CDS             complement(join(44143. .44452,45240. .45558,45719. .45878))
                /gene="F13M22.14"
                /codon_start=1
                /product="putative expansin protein"
                /db_xref="PID:g3236246"
                /translation="MTATAFRVGLWLAVTASFLLTATNAKIPGVYSGGPWQNAHATFY
                GGSDASCTMGGACGYGNLYSQGYGVNTAALSTALFNGFSCGACFEIKCTDDPRWCVP
                GNPSILVTATNFCPPNFAQPSDDGGWCNPPREHFDLAMPMFLKIGLYRAGIVPVSYRR
                VPCRKIGIRFTVNGFRYFNLVLVTNVAGAGDINGVSVKGSKTDWVRMSRNWGQNWQS
                NAVLIQOSLSFVTASDRRSSTSWNVAPATWQFGQTFSGKNFRV"

repeat_region   44795. .44903
                /rpt_family="AT_rich"

repeat_region   46287. .46312
                /rpt_family="AT_rich"

misc_feature    complement(47301. .47444)
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"

repeat_region   48559. .48589
                /rpt_family="AT_rich"

mRNA            <48689. .>50845
                /gene="F13M22.15"

gene            <48689. .>50845
                /gene="F13M22.15"

CDS             48689. .50845
                /gene="F13M22.15"
                /codon_start=1
                /product="SCARECROW-like protein"
                /db_xref="PID:g3236247"
                /translation="MITEPSLTGISGMVNRNRLSGLPDQPSSHSFTPVTLYDGFNYNL
                SSDHINTVVAAPENSVFIREEEEDPADDFDFSDAVLGYISQMLNEEDMDDKVCMLQ
                ESLDLEAAERSLYEAIGKKYPPSPERNLAFAERNSENLDRVVPGNYTGGDCIGFGNGG
                IKPLSSGFTLDFRNPQSCSSILSVPQSNGLITIYGDGIDESSKNNRENHQSVWLFRRE
                IEEANRFNPEENELIVNFREENCVSKARKNSSRDEICVEEERSSKLPAVFGEDILRSD
                VVDKILVHVPGGESMKEFNALRDVLKKGVEKKKASDAQGGKRRARGRGRGRGGGGGG
                QNGKKEVDVLRSLLHCAQAVAADDRRCAGQLLKQIRLHSTPFGDGNQRLAHCFANGL
                EARLAGTGSQIYKGIVSKPSAAAVLKAHQLFLACCPFRKLSYFIINKTIRDLVGNSQ
                RVHVIDFGILYGFQWPTLIHRFSMYGSPKVRITGIEFPQPGFRPAQRVEETGQRLAAY
                AKLFGVPFEYKAIAKKWDAIQLEDLDIRDEITVVNCLYRAENLHDESVKVESCRDTV
                LNLIGKINDPLFVFGIVNGAYNAPFFVTRFREALFHSSSIFDMLETIVRDEERMFL
                EMEVFGREALNVIACEGCWERVERPETYKQWHVRAMRSGLVQVPFDPSIMKTSLHKVHT
                FYHKDFVIDQDNRWLLQGWKGRTVMALSVWKPESKA"

repeat_region   49591. .49709
                /rpt_family="(GAA)n"

repeat_region   49635. .49689
                /rpt_family="(GGA)n"

misc_feature    51112. .51155
                /note="exon predicted by xgrail, quality good"

misc_feature    complement(51301. .51387)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"

mRNA            complement(join(51428. .51666,51782. .51905,51988. .52049,
                52131. .52264,52306. .52342,52427. .52698,52779. .>53043))
                /gene="F13M22.16"

gene            complement(51428. .>53043)
                /gene="F13M22.16"

CDS             complement(join(51547. .51666,51782. .51905,51988. .52049,
                52131. .52264,52306. .52342,52427. .52698,52779. .53043))
                /gene="F13M22.16"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236248"
                /translation="MAMMTTTTTFFHPLLPANTYKSGAVASSFVSVPRSSSLQFRSL
                VSDSTSICGPSKFTGKNRRVSVTVSAAATTEPLTVLVTGAGGRTGQIVYKKLKERSEQ
                FVARGLVRTKDSKEKINGEDEVFIGDIRDTASIAPAVEGIDALVILTSAVPQMKPGFD
                PSKGGRPEFFFDDGAYPEQVDWIGQKNQIADGVDDGLSFVVTARAAGVKQIVLVGSM
                GGTNINPLNSIGNANLVWKRKAEQYLADSGIPYTIIRAGLQDKDGGIRELLVGSM
                DELLETERTRIARADVAEVCVQALQEEAKFKALDLASKPEGTGTPTKDFKALFTQVT
                TKF"

mRNA            complement(join(<53450. .54129,54196. .54289,54371. .54453,
                54545. .54596,54711. .54868,54958. .>56602))
                /gene="F13M22.17"

gene            complement(<53450. .>56602)
                /gene="F13M22.17"
                /note="predicted by grail"

CDS             complement(join(53450. .54129,54196. .54289,54371. .54453,
                54545. .54596,54711. .54868,54958. .56602))
                /gene="F13M22.17"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236249"
                /translation="MEDGNKLDRKKTMTMNWEGLGFDEDEDDDRFFETHDRLSSALTF
                DMANASSDEDEDFDDCRLSFSSAVSSLTTASRKFRTPAMSPDYDIWMAAPGSISERR
                RRLLHGMGLASNKDMVSAVSIRRVVSNAPVVSNGEEKKMKKKMNHEVDEQDHDHVPVM
                LARSRESDIERFLIEKRRKEEILGKISKQRLTRTYSTICPTRTRICQYQTPIRQTPA
                VCGNGKALRGGGEALTVMSNARVGAFFLIKNLDTGEFIVNEYDEDGMWNRLSDLQT
                GKQLTEEFEKCVGYSPVVKELMRRENVNRINYEPLMDLRKFNSYLSKSVRLSKRRGA
                ALLKNIKGVAHSMSLRVADKDVSDGSTDSPKKGKDHKHGKANEWVKVRPTGKSYKELS
                ALHMCQEIQAHEGAVWTIKFSODAHYLASGGADRVIHVWECCELMSMNEGSLTPIH
                PSLCDSSGNEITVEKKKKGGSSGRRHNHIPDYVHVPETVFSFSDKPVCSLKGHLDA
                IIDLSWKSKQLLLSSSMDKTVRLWDIETKTCLKLFAHNDYVTCIQFSPVDENYFLSGS
                LDAKIRIWSIQDRHVVEWSDLHEMVTAACYTPDGQGALIGSHKGICRAYDTEDCKLSQ
                TNQIDVQSNKKSQAKRKITSFQFSPVNPSEVLVTSADSRIRILDGSEVIHKFKGFRNT
                CSQLSASYSQDGKYIICASEDSQVYLWKNDFHRTRSTLTTQSHEHFHCKDVSAAVPWH
                GHVRGEPPVQICHSKRHSKRISTSSQPSTISSPTKETSATGPTTSNRNKKSGLPPM
                PKKAATKSQIQPEEEAGPELGSSESFRSSMNSSEQHSSRFGESPSINTSSRLSSWSWF
                DSGGHGPQTIQPTAWGMVIVTATVHGEIRSYQNFGLPRRIGRQTTLF"

repeat_region   complement(56145. .56194)
                /rpt_family="(GAA)n"

misc_feature    56686. .56744
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"

misc_feature    complement(56743. .56808)
                /note="exon predicted by xgrail, quality marginal"

misc_feature    complement(57092. .57128)
                /note="exon predicted by xgrail, quality marginal"

misc_feature    complement(57429. .57497)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"

mRNA            complement(57759. .57834)
                /note="exon predicted by xgrail, quality excellent"
                complement(join(<57887. .58388,59719. .59846,60078. .60132,
                60218. .60291,60384. .60527,61543. .>61581))
                /gene="F13M22.18"

gene            complement(<57887. .>61581)
                /note="predicted by genefinder"

CDS             complement(join(57887. .58388,59719. .59846,60078. .60132,
                60218. .60291,60384. .60527,61543. .61581))
                /gene="F13M22.18"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236250"
                /translation="MPVRVVESNTPAQAFSGTQNVSNQQKEEAWRVNVQIGGIDLEHG
                YLCGTMEALNVPMADTPVITFWEGEIVDGKNTFYTGKWEATREDDMRHWSKFPSFSP
                LQRWKEQYFVNVGTDCGLTIAGFYVCFSCSDGSISGFYDPNSSFHHMISSSKNVL
                KMEEVEVSKKRKFQTDQSDELSLLPLSKHTCFANVACSENTNGNSEIDTEYSMSSYVN
                STTSMECNDIEMKEESSGSCGEDKMISFESHLDYIYGTQNLEDFSEKVIENILYLDE
                QEEEEDAKGCSSNAAKFVLSSGRWTVNQGKENRILV"

repeat_region   58770. .58795
                /rpt_family="(GA)n"

repeat_region   59165. .59234
                /rpt_family="AT_rich"

repeat_region   59249. .59312
                /rpt_family="AT_rich"

repeat_region   complement(59446. .59478)
                /rpt_family="AT_rich"

misc_feature    complement(61713. .61851)
                /note="exon predicted by xgrail, quality good_shadowexon"

mRNA            join(<62177. .62406,62814. .63035,63131. .63244,63337. .63399,
                63504. .63642,63940. .63995,64072. .64145,64247. .64394,
                64470. .64652,64757. .64844,64948. .65058,65257. .65354,
```

gene
```
                65591..65708,65788..65882,66002..66108,66213..>66306)
                /gene="F13M22.19"
                <62177..>66306
                /gene="F13M22.19"
```

CDS
```
                join(62177..62406,62814..63035,63131..63242,63337..63399,
                63504..63642,63940..63995,64072..64145,64247..64394,
                64652,64757..64884,64948..65058,65255..65354,
                65591..65708,65788..65882,66002..66108,66213..66306)
                /gene="F13M22.19"
                /codon_start=1
                /product="putative phosphoribosylaminoimidazole
                carboxylase"
                /db_xref="PID:g3236251"
                /translation="MLLKQSSAAVLVVGNTPVLHTSRSTYRVGPFPVTRTQSFQSL
                TMANLQKLPTSSSGKLNTASAVPCCSHDASPISENRENKHVHGVSEKIVGVLGGGQLG
                RMLCQAASQLAIKVMILDPSKNCSASALSYGHMVDSFDDSATVEEFAKRCGVLTVEIE
                HVDVDTLEKLEKQGVDCQPKASTIRIIQDKVMQKVHFSQHGIPLPEFMEISDIEGARK
                AGELFGYPLMIKSKRLAYDGRGNAVANNQDELSSAVTALGGFSRGLYIEKWAPFVKEL
                AVIVARGRDGSMVCYPVVETIHRDNICHIVKAPADVPWKINKLATDVAQKAVGSLEGA
                GVFAVELFLTEDSQILLNEVAPRPHNSGHQTIECCVTSQFEQHLRAVVGLPLGDPSMR
                TPASIMYNILGEDDVIDGEAGFKLAHRLIARALCIPGASVHWYDKPEMRKQRKMGHIT
                LVGQSMGLIEQRLQCILSEQSHQVHETPRVAIIMGSDTDLPVMKDAAKILLDLFGVTHE
                VKIVSAHRTPEMMYTATSAHSRGVQVIIAGAGAAHLPGMVASLTLPLPVIGVPVRAT
                RLDGVDSLLSIVQMPRGVPVATVAINNATNAALLAVRMLGISDTDLVSRMRQYQEDMR
                DENLNKGEKLETEGWESYLNQ"
```

mRNA
```
                complement(join(<66647..66829,67056..67229,67307..67405,
                67776..67976,68080..68187,68273..68651,68917..69175,
                69620..70064,70183..>70242))
                /gene="F13M22.20"
```

gene
```
                complement(<66647..>70242)
                /gene="F13M22.20"
```

CDS
```
                complement(join(66647..66829,67056..67229,67307..67405,
                67776..67976,68080..68187,68273..68651,68917..69175,
                69620..70064,70183..70242))
                /gene="F13M22.20"
                /codon_start=1
                /product="CER1-like protein"
                /db_xref="PID:g3236252"
                /translation="MASRPGFLTDWPWTPLGSFKYLLLAPLVFDSIYSYATIRDHEKL
                LIVAVTWRIVHSQIWISLSRYQTAKGTKRILNKSIEFDQVDRERTWDDQIIFNTLIV
                YLTKVYVSGTSTIPFWRTDGVILVALLHAGPVEFIYYWFHRALHHHFLYSRYHSHHHS
                SIVTEPITLCATNSKPWVLIVAVVHPFAEHIGYTLLGLPLITTFMCGTVSVVSIALY
                LTYIDFMNNMGHCNFELIPKFLFSLLPPLKFLCYTPSFHSLHHTQFRTNYSLFMPMYD
                YIYGTTDECSDSLYETSLEKEEEKPDAIHLTHLTSLDSIYHLRLGFASLSSHPLSSRC
                YLFLMKPFALILSFILRSFSFQTFVVERNRFRDLTLHSHLLPKFSSHYMSHQQKECIN
                KMIEAAILEADKKGVKVMSLGLLNQGEELNGYGEMVVRRHPKLKIRIVDGGSLAAEVV
                LHSIPVGTKEVLFRGQITKVARAIVFSLCQNAIKVMVLRKEEHSMLAEFLDDKCKENL
                VLTTNYYPMIWLVGDGLSTKEQKMAKDGTLFLPFSQFPPKTLRKDCFYHTTPAMIIPH
                SAQNIDSCENMLGRRVMSAWRVGGIVHALEGWKEHECGLDDNSIINPPRWEAALRNG
                FQPLLLPSLET"
```

repeat_region
```
                complement(67680..67715)
                /rpt_family="AT_rich"
```

misc_feature
```
                complement(70565..70634)
                /note="exon predicted by xgrail, quality marginal"
```

mRNA
```
                complement(70808..73053)
                /gene="F13M22.21"
```

gene
```
                complement(70808..73053)
                /gene="F13M22.21"
                /note="highly similar to GB:X95909(polymorphism exists
                between nucleotides 73005 and 72085 which is a GA repeat.
                We found 6 copies in our sequence whereas only 5 copies
                exist in GB:X95909)"
```

CDS
```
                complement(71000..73027)
                /gene="F13M22.21"
                /codon_start=1
                /product="receptor-like protein kinase"
                /db_xref="PID:g3236253"
                /translation="MFLKLLTIFFFFFFNLIFQSSSQSLNFAYNNGFNPPTDLSIQGI
                TTVTPNGLLKLTNTTVQKTGHAFYTKPIRFKDSPNGTVSSFSTSFVFAIHSQIALSG
                HGIAFVVAPNASLPYGNPSQYIGLFNLANNGNETNHVFAVELDTILSTEFNDTNDNHV
                GIDINSLKSVQSSPAGVWDEKGQFKNLTLISKKPMQVWVDYDGRTNKIDVTMAPFNED
                KPTRPLVTAVRDLSSVLLQDMVGFSSATGSVLSEHYILGWSFGLNEKAPPLALSRLP
                KLPRFEPKRISEFYKIGMPLISLFLIFSFIFLVCYIVRRRRKFAEELEEWEKEFGKNR
```

```
                FRFKDLIYATKGFKEKGLLGTGFGSVYKGVMPGTKLEIAVKRVSHESRQGMKEFVAE
                IVSIGRMSHRNLVPLLGYCRRGELLLVYDYMPNGSLDKYLYNTPEVTLNWKQRIKVI
                LGVASGLFYLHEEWEQVVIHRDVKASNVLLDGELNGRLGDFGLARLYDHGSDPQTTHV
                VGTLGYLAPEHTRTGRATMATDVFAFGAFLLEVACGRRPIEFQOETDETFLLVDWFG
                LWNKGDILAAKDPNMGSECDEKEVEMVLKLGLLCSHSDPRARPSMRQVLHYLRGDAKL
                PELSPLDLSGSGMFGVHDGFSELGMSYSSSVFKGFTGGSSIADSQLSGGR"
```

repeat_region
```
                72962..73005
                /rpt_family="(GAA)n"
```

repeat_region
```
                complement(73447..73473)
                /rpt_family="AT_rich"
```

mRNA
```
                join(<74148..74206,74511..74566,75952..>77285)
                /gene="F13M22.22"
```

gene
```
                <74148..>77285
                /gene="F13M22.22"
```

CDS
```
                join(74148..74206,74511..74566,75952..77285)
                /gene="F13M22.22"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236254"
                /translation="MVQIVPHVANAHMVKTSRKGLLALVVGVEECCPTRRCDLRDPN
                LVVDSNCLLSRSTHQSSLVEDSIIINHFRILNFNNRASIFLCLRIGVICSTQSKMYGG
                KSYILKGSVLSLELILLLVSAKNQPRISLIEVKTCNLAKGEWVEDKRPLYS
                GFECKQWLSNIFSCRVMGRPDFSFEGYRWQPEGCNIPEFNRVNFLRRMQNKTIAFIGD
                SLGREQFQSLMCMATGGKESPEVQNVGSEYGLVIPKGAPRPGGWAYRFPTTNTTVLSY
                WSASLTDLVPMNNTDPPHLIAMHLDRPPAFIRNYLHRFHVLVNTGHHNRPHEKCN
                WVMHVNGTRVEGGYFKNVENAKIFTIHSLVKWLDAQLPLHPRLKAFFTTISPRHEKCN
                NTIPLSRGSKITGEGGSLDTIVESAVNGTRVKILDITALSKLRDEAHIAGCKLKPKKA
                SNVTSAPFNDCLHWCLPGIPDTWNELLIAQL"
```

repeat_region
```
                complement(75409..75618)
                /rpt_family="AT_rich"
```

repeat_region
```
                complement(77388..77424)
                /rpt_family="AT_rich"
```

repeat_region
```
                complement(77464..77524)
                /rpt_family="AT_rich"
```

misc_feature
```
                77550..77635
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
```

mRNA
```
                complement(join(<78252..78416,78604..79152,79555..79945,
                80111..>80604))
                /gene="F13M22.23"
```

gene
```
                complement(<78252..>80604)
                /gene="F13M22.23"
                /note="predicted by genscan"
```

CDS
```
                complement(join(78252..78416,78604..79152,79555..79945,
                80111..80604))
                /gene="F13M22.23"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236255"
                /translation="MAIFDFKLHHKPPSIFPATTVKPSHVLSLVAKFFTICIFISV
                AMISFYIIFFGCSDCHITGLRLRGPENVNLTNSSASSTIRRPNQSSEATDISHIA
                FGIGGSIOTWRDRSRYSELWWRPNVTRGFIWLDEEPPLNMTWLSTSPPYQVSADTSRF
                SYTCWYGSRSAIRMARIIKETFELGLTDVRWFIMGDDDTVFFVDNLITVLNKIDHNQM
                YYIGGNSEVEDDIVHSYAMAYGGGGIAISYPLAVELVKLLDGCIDRYASLYGSDQKI
                EACLSEIGVPLFKELGFHQVDIRGNPYGLLAAHPVAPLVTLHHLDYVDPFPGCTQID
                ALRRLVSAYKTDFSRIIOHSFCHDQTRNWYVSVSWGYTIQIYPLVTAKELETPFLTF
                KSWRTSSSEPFSPDTRPISEDPCERPLVYFLDRVYEVGSGQTLTTYRKHVEVGESTQC
                NSPDYSRANPVEFIDVSTTLTPDLWKMAPRRQCCEIVNSEEDSESVINKIRHFNPL
                ESVTPQSKHPNDEWTDSVLSTETI"
```

repeat_region
```
                complement(79324..79361)
                /rpt_family="AT_rich"
```

repeat_region
```
                80698..80727
                /rpt_family="AT_rich"
```

misc_feature
```
                complement(81151..81260)
                /note="exon predicted by xgrail, quality excellent"
```

misc_feature
```
                82037..82089
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
```

misc_feature
```
                82508..82558
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
```

mRNA
```
                complement(<83772..>84686)
                /gene="F13M22.24"
```

gene   complement(<83772..>84686)
                  /gene="F13M22.24"
CDS   complement(83772..84686)
                  /gene="F13M22.24"
                  /codon_start=1
                  /product="putative zinc finger protein"
                  /db_xref="PID:g3236256"
                  /translation="MEXPGGFWIPKKSNKESSWEELAFAEDDAAGSLWPPRSYTCSFC
                  RREFKSAQALGGHMNVHRDRARLKQADDQYLFKSSSSPEYPSHKDSDNIHETSCYT
                  LVFNTKPNYFKTQHSCVIDLSSSSSLPYLPSRVSSGLPGKQHTSSSPSFVVEPSKN
                  SKYIPSSSPWSCPSTVVEQKSCDLYEIPAMGEKKRKTESDVPKIGHKAKLSLGNTD
                  NEESKHKGDRREDLDELRLGTDPPKGI"
repeat_region   84230..84321
                  /rpt_family="(GAA)n"
repeat_region   84910..84962
                  /rpt_family="AT_rich"
repeat_region   85196..85220
                  /rpt_family="(TA)n"
repeat_region   complement(85246..85307)
                  /rpt_family="AT_rich"
misc_feature   complement(85742..85810)
                  /note="exon predicted by xgrail, quality marginal"
repeat_region   86102..86122
                  /rpt_family="AT_rich"
misc_feature   86195..86308
                  /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   86660..86687
                  /rpt_family="AT_rich"
misc_feature   86972..87002
                  /note="exon predicted by xgrail, quality marginal"
mRNA   87049..>87569
gene   87049..>87569
                  /gene="F13M22.25"
CDS   87120..87332
                  /gene="F13M22.25"
                  /note="unknown protein"
                  /codon_start=1
                  /db_xref="PID:g3236257"
                  /translation="MQIRYGNESGQQLSSSYYVRLPGDSGRFVRSDLDVSGPGSNSSNR
                  ASSSTSTTTFAVYAGVQSPVSRQVVT"
mRNA   join(<88062..88169,88258..88308,88608..88684,88770..88857,
                  88934..89209,89301..>89585)
gene   <88062..>89585
                  /gene="F13M22.26"
CDS   join(88062..88169,88258..88308,88608..88684,88770..88857,
                  88934..89209,89301..89585)
                  /gene="F13M22.26"
                  /codon_start=1
                  /product="putative alcohol dehydrogenase"
                  /db_xref="PID:g3236258"
                  /translation="MAAPIRFFELNTGAKLPCVGLGTYAMVATAIEQAIKIGYRHIDC
                  ASIYGNEKEIGGVLKKLIGDGFVKREELFITSKLWSNDHLPEDVPKALEKTLQDLQID
                  YVDLYLIHWPASLKKESLMPTPEMUIKPDITSTWKAMEALYDSGKARAIGVSNFSKK
                  QNPIVTEVAEKLGKTTAQVALRWGLQTGHSVLPKSSSGARLKENLDVFDWSIPEDLFT
                  KFSNIPQASSFIFLFFSR"
misc_feature   89723..89798
                  /note="exon predicted by xgrail, quality excellent"
misc_feature   complement(89743..89842)
                  /note="exon predicted by xgrail, quality
                  excellent_shadowexon"
repeat_region   complement(90572..90624)
                  /rpt_family="AT_rich"
mRNA   join(<90955..91074,91214..91264,91485..91561,91649..91736,
                  91942..92291,92217,92291..>92551)
gene   <90955..>92551
                  /gene="F13M22.27"
CDS   join(90955..91074,91214..91264,91485..91561,91649..91736,
                  91942..92217,92291..92551)

                               repeat_region   /gene="F13M22.27"
                                                 /codon_start=1
                                                 /product="putative alcohol dehydrogenase"
                                                 /db_xref="PID:g3236259"
                                                 /translation="MANAITTFFKLNTGAKFPSVGLGTWQASPGLVGDAVAAVKIGYh
                                                 HIDCAQIYGNEKEIGAVLKKLFEDRVVKREDLFITSKLWCTDHDPQDVPEALNRTLKc
                                                 LQLEYDLYLIHWPARIKKGSVGIKPENLLPVDIFSTWKAMEALYDSGKARAIGVSNF
                                                 STKKLADLLELARVPPAVNQVECHPSWRQTKLQEFCKSKGVHLSAYSPLGSPGTTWLK
                                                 SDVLKNPILNMVAEKLGKSPAQVALRWGLQMGHSVLPKSTNEGRIKENFNVFDWSIPD
                                                 YMFAKFAEIEQVRS"
                               repeat_region   complement(91371..91445)
                                                 /rpt_family="(CAAAA)n"
                               misc_feature   92640..92726
                                                 /note="exon predicted by xgrail, quality excellent"
                               misc_feature   complement(93519..93682)
                                                 /note="exon predicted by xgrail, quality good_shadowexon"
                               misc_feature   93640..93819
                                                 /note="exon predicted by xgrail, quality excellent"
                               repeat_region   complement(93688..93779)
                                                 /rpt_family="(CAT)n"
                               BASE COUNT   29945 a 17475 c 16793 g 29632 t
                               ORIGIN

                               Query Match   7.1%;   Score 70.4;   DB 9;   Length 93845;
                               Best Local Similarity   74.2%;   Pred. No. 4.9e-05;
                               Matches 89; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

                               Qy   363 attctcttggcctccaagatcctatacatgtagcttttgtaaaggaattagatctgc 422
                                         || || |||||| |||||||| ||||||| ||||| |||||||||| |||| ||||| ||
                               Db 84594 ATCTTTGTGGCCGCCAAGATCTTATACGTGTAGCTTCTGCCCGGAGAGAATTAAATCGGC 84535

                               Qy   423 tcaagtcttgtggacacatgaatgttcatagaagagatagagccatttgagacaatc 482
                                         |||||| || |||| |||| |||||||| ||||| |||||| |||||||| ||||| |||
                               Db 84534 TCAAGCCCTCGTGGTGGCCATATGAACGTTCACCGAAGACAGCAGCACGGCTTAAGCAAGC 84475

                               RESULT 5
                               ATAC004684/C
                               LOCUS   ATAC004684   93845 bp   DNA   PLN   17-JUN-1998
                               DEFINITION Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence,
                                           complete sequence.
                               ACCESSION   AC004684
                               NID   g3236234
                               KEYWORDS   HTG.
                               SOURCE   thale cress.
                               ORGANISM   Arabidopsis thaliana
                                           Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
                                           Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
                                           eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
                               REFERENCE   1 (bases 1 to 93845)
                               AUTHORS   Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,
                                           Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
                                           Somerville,C.R. and Venter,J.C.
                               TITLE   Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence
                               JOURNAL   Unpublished
                               REFERENCE   2 (bases 1 to 93845)
                               AUTHORS   Rounsley,S.D. and Lin,X.
                               TITLE   Direct Submission
                               JOURNAL   Submitted (14-MAY-1998) The Institute for Genomic Research, 9712
                                           Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
                               REFERENCE   3 (bases 1 to 93845)
                               AUTHORS   Rounsley,S.D.
                               TITLE   Direct Submission
                               JOURNAL   Submitted (17-JUN-1998) The Institute for Genomic Research, 9712
                                           Medical Center Dr., Rockville. MD 20850, USA
                               COMMENT   On Jun 17, 1998 this sequence version replaced gi:3172168.
                                           Address all correspondence to:
                                           Steve Rounsley
                                           The Institute for Genomic Research
                                           9712 Medical Center Dr,
                                           Rockville, MD 20850,
                                           USA

```
                     e-mail: rounsley@tigr.org
                     BAC clone F13M22 is from Arabidopsis chromosome II and is near the
                     molecular marker ve018.
                     The orientation of the sequence is from SP6 to T7 end of the BAC
                     clone.
                     Genes were identified by a combination of three methods: Gene
                     prediction programs including GRAIL (available by anonymous ftp
                     from arthur.epm.ornl.gov), Genfinder (Phil Green, University of
                     Washington), Genscan (Chris Burge,
                     http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
                     (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
                     complete sequence against a peptide database and the Arabidopsis
                     EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
                     Annotated genes are named to indicate the level of evidence for
                     their annotation. Genes with similarity to other proteins are named
                     after the database hits. Genes without significant peptide
                     similarity but with EST similarity are named as 'unknown' proteins.
                     Genes without protein or EST similarity, that are predicted by more
                     than two gene prediction programs over most of their length are
                     annotated as 'hypothetical' proteins. Genes encoding tRNAs are
                     predicted by tRNAscan-SE (Sean Eddy,
                     http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
                     identified by repeatmasker (Arian Smit,
                     http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
                     genomic sequence that are not annotated as genes but have predicted
                     exons by GRAIL are annotated as misc features.
FEATURES             Location/Qualifiers
     source          1..9385
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="ve018"
                     /clone="F13M22"
     misc_feature    complement(10..29)
     misc_feature    168..249
                     /note="exon predicted by xgrail, quality marginal"
     mRNA            complement(464..>619
                     /gene="F13M22.1"
     gene            <464..>619
                     /note="predicted by genscan"
                     /gene="F13M22.1"
     CDS             <464..619
                     /gene="F13M22.1"
                     /note="hypothetical protein, 5' partial"
                     /codon_start=1
                     /db_xref="PID:g3236260"
                     /translation="SRGDVRVEQRRTHKDGEDKRLLDPATRRKLETNAIIASGTKGF
                     EIGNGNL"
     mRNA            complement(join(835..1212,1373..1513,1600..1693,
                     1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
                     2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
                     4167..4353,4486..4549,4633..4842,4938..5006,5088..>5654))
                     /gene="F13M22.2"
     gene            complement(835..>5654)
                     /gene="F13M22.2"
     CDS             complement(join(1099..1212,1373..1513,1600..1693,
                     1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
                     2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
                     4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
                     /gene="F13M22.2"
                     /note="unknown protein"
                     /codon_start=1
                     /db_xref="PID:g3236235"
                     /translation="MPKEENGQLKRDRLDDDTDEGNKGDHFPSKKQAKEASNDDITSE
                     ISNPVASPVESTSLFRDVSSQPVKSGLVECSGSDFGSEETVSDDASVVGSSQTEQSSD
                     VLPSRFVLEIPKHLSSTGTKTIFFKLSKPKKEFDLPLIKDHTWDAGYVKMPKKKIVS
                     LSYPSNVKKLLETGILEGARVKYISTPPVRQLLGIIHSGGYLCGCTCNFSKVLSAYE
                     FEQHAGAKTRHPNNHIFLENRRAVYNIVQELKTAPRVVLEEVIRNVAGSALNEEGLRA
                     WKGDGQVSNRVIACLTGILRIIPLSVNALLSYHFVLLSETDSCESCSYLGPGLDES
                     QSLTPCSVENHYFSEKTYAKDTLDEPKRIAKKLTSHVSGTGCHKKVSEGSNRKRDNDL
                     HRLLFMPNGLPDGTELAYYVKTQISPSQFEAHAGMAARRQPYRHIFISSGLSLHDiAM.
                     SLANGHVITTGDSDDMCSICGDGDLLLCAGCPQAFHTACLKFQSMPEGTWYCSSCND
                     GPISSKATTTDPSGNARPIVIRLSRVVKAPESDIGGCVFCRSHDFSIGKFDDRTVIL
                     CDQCEKEYHVGCLRENGFCDLKEIPQKSKRYPEHLPLLSRAAVIFRECFDIVAKSGR
                     DMICRKDREKGIFTDIGDIVEWRILSGKSRYPEHLPLLSRAAVIFRECFDIVAKSGR
                     DLIPVMYGRNISGQEFGGMYCLVLIVNSLVVSAALLRIFGQEVAELPIVATSREYQG
                     RGYFQGLYACVENLLSSLNVENLVLPAEEAESIWTKKFGFTKMSDQQLQEYQKEVQL
                     TIFKGTSMLEKKVPKATTGLSESTLI"
     repeat_region   6010..6135
                     /rpt_family="(GGAA)n"
     misc_feature    6023..6202
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   6179..6236
                     /rpt_family="POLY_A"
     mRNA            join(<647..6702,7161..7304,7322..>7499)
                     /gene="F13M22.3"
     gene            <647..>7499
                     /note="predicted by genscan"
                     /gene="F13M22.3"
     CDS             join(6647..6702,7161..7304,7322..7499)
                     /gene="F13M22.3"
                     /note="hypothetical protein"
                     /codon_start=1
                     /db_xref="PID:g3236236"
                     /translation="MTTWKKVREENKSFOFLFKFAWRIPHFVYGATWTLFLTLTVAII
                     SLAPEFAFVSAIFPSSSEVFSRRHIGSAGSLGGLVFTGESFPEVEDGPRCAAGFCSN
                     CCGSFRRRKKDYGPLGGRSPLI"
     mRNA            complement(join(<7761..7931,8019..8065,8154..8238,
                     8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
                     /gene="F13M22.4"
     gene            complement(<7761..>9886)
                     /gene="F13M22.4"
     CDS             complement(join(7761..7931,8019..8065,8154..8238,
                     8323..8401,8494..8647,8824..8903,9001..9232,9769..9886))
                     /gene="F13M22.4"
                     /codon_start=1
                     /product="putative ribotol dehydrogenase"
                     /db_xref="PID:g3236237"
                     /translation="MGIYGVMTGKKGKSGFGSASTAEDVTQAIDASHTAIITGGTSG
                     IGLEAARVLAMRGAHVIIARNPKAANESKEMILQNPNARVDYLQIDVSSIKSVRSF
                     VDQFLALNVPLNILINNAGVMFCPFKLTEDGIESQFATNHIGHFLLTNLLLDKMKSTA
                     RESGVQGRIVNLSSIAHYTTYSEGIKFQGINDPAGYSERRAYGQSKLSNLLHSNALSR
                     RLQEEGVNTTINSVHPGLVTTNLFRYSGFSMKVRAMTFLFWKNIPQGAATTCYVALii
                     PDLEGVTGKYFGDCNIVAPSKFATNNSLADKLWDFSVFLIDSISK"
     repeat_region   8658..8684
                     /rpt_family="AT_rich"
     repeat_region   complement(9842..9915)
                     /rpt_family="(GA)n"
     misc_feature    10503..10670
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   10725..10746
                     /rpt_family="AT_rich"
     mRNA            complement(join(11416..11526,11608..12116,12551..12628,
                     12737..13604))
                     /gene="F13M22.5"
     gene            complement(11416..13604)
                     /gene="F13M22.5"
     CDS             complement(join(11610..12116,12551..12628,12737..13522))
                     /gene="F13M22.5"
                     /codon_start=1
                     /product="putative ARF1 GTPase activating protein"
                     /db_xref="PID:g3236238"
                     /translation="MAAARRLRTLQSQPENKVCVDCSQKNPQWASISYGIFMCLECSG
                     KHRGLGVHISFVRSVTMDSWSEIQIKMDAGGNERLNNFLAQYGISKETDIISKYNSN
                     AASVYRDRIQALAEGRQWRDPIVKESVGGGLMNKKPPLSQGGRDSGNGGWDNWDND
                     DSFRSTDMRRNQSAGDFRSSGGRGAPAKSKSSEDIYSRSQLEASAANKESFFAKRMAE
                     NESKPEGLPPSQGGYVGFGSSPGPAPRSNQQSGGGDVFSVMSEGFGRLSLVAASAAN
                     VVQTGTMEFTSKVKEGGLDQTVSETVNVVASKTEIGQRTWGIHKGVVAIASQKVEEF
                     TKEEASTWNQQNKTEGNGYYQNSGIGNKTANSSFGGSGSQSSSGHNNSYRNSNSWDDWG
                     EENNSKKEAAPKVSTNDDDGGWAGWDDNDAKDDDFYYQPASDKKSVGHNGKSDTAW
                     TGGGFL"
     repeat_region   11697..11757
```

misc_feature   /rpt_family="(CAT)n"
13716. .13952
/note="exon predicted by xgrail, quality
excellent_shadowexon"

misc_feature   complement(14172. .14309)
/note="exon predicted by xgrail, quality marginal"

mRNA   join(14707. .14491,15566. .15765,15858. .15964,16057. .16189,
16311. .16602,16755. .17294)
/gene="F13M22.6"

gene   14707. .17294
/note="identical to GB:U40269"
/gene="F13M22.6"

CDS   join(15634. .15765,15858. .15964,16057. .16189,16311. .16602,
16755. .17182)
/gene="F13M22.6"
/codon_start=1
/product="origin recognition complex protein"
/db_xref="PID:g3236239"
/translation="MEDIENIEEDEYGFSRNYFLAKELGGASKRSAHKLSDIHIVDEQ
ELRETASTIEMKHSKEISELMSDYKTMYSKWVFELRCGFGLLMYGFGSKKALVEDFAS
ASLTDYSVVVINGYLPSVNLKQVLLALAELLSELIKCKRKSSGSLSKGQETFPSRSMD
DILSFLHGPQSGDKDCFICVVVHNIDGPALRDPESOQTLARLSSCSHIRLVASIDHVN
APLLWDKKMVHKQFNWLWHHVPTFAPYNVEGVFFPLVLAQGSTAQTAKTAAIVLQSLT
TKRNSDGQECLNIPLTSDAIRQLLLDLNQ"

mRNA   complement(<17532. .>19016)
/gene="F13M22.7"

gene   complement(<17532. .>19016)
/gene="F13M22.7"

CDS   complement(17532. .19016)
/gene="F13M22.7"
/note="unknown protein"
/codon_start=1
/db_xref="PID:g3236240"
/translation="MENHHPSTLLSMDSSASHEELDLEMNNNRQSLLSGPPDINLPL
SAERSPPPPWNLDACDILDVGLGSQAYETENYMSVVPKVGRKCAKRVDSIWGAWFFF
SFYFKPALNEKSKAKIVRDSNGISGFDKSDLKLDVFLVQHDMENMYMWVFKERPENAL
GKMQLRSYMNGHSRQGDRLFPFSVEKGFVRSHRMQRKHYRGLSNPQCVHGIELVPLPN
LTCLDEEERKRWMELTGRDLNFTIPPEASDFGSWRNLPNTDFELERPTPSLKNPTANH
SKKLLNGSGLNLSTOPSNHSNGEATDLSPSSHKKRKDLFSNGIHEEECCLTVNPQPPV
IEAHQNELPTWSNEFTGAMKNVYGPVTAAKTIYEDEEGYLIIISLPFVDLNSVKVSWR
NTLTHGIIKVSCLSTSRVPFIKRHDRTFKLTDSASEHCCPPGEFVREIPLSNRIPEDAN
IEAYYDGPGSVLEILVPKLRAGPEEHEVRVCLRPNLGGNDLMLT"

misc_feature   complement(19045. .19091)
/note="exon predicted by xgrail, quality marginal"

misc_feature   19455. .19545
/note="exon predicted by xgrail, quality
excellent_shadowexon"

mRNA   join(<20811. .21507,21667. .21805,22028. .22434,22931. .23006,
23118. .23233,23331. .>23788)
/gene="F13M22.8"

gene   <20811. .>23788
/gene="F13M22.8"

CDS   join(20811. .21507,21667. .21805,22028. .22434,22931. .23006,
23118. .23233,23331. .23788)
/gene="F13M22.8"
/codon_start=1
/product="putative zinc finger protein"
/db_xref="PID:g3236261"
/translation="MFNNTTTSFGSGPGIVVVPTPATTVPTTDPPGTTITSNSTFIII
GPPPPPAPPRSIDLTPLKLIFVVIAFVAVPALVVALFNGPCSSSRRNSSSSRTSSS
SDDTPHATVDFTPIPITETTVTSESGGKFKDPTHSKEIGNECSVCLMVFTDSDELRQLSE
CKHAFHVLCIETWLKDHPNCPICRTDVSVKQQTEAPNVPVNVNGNVNRSGGNRRVSAT
SRDDDWRQGLPDASTTVSVSVRDVKSDVDSQQAHPTVYEDLILIGRLLIHTRLPIRTL
SFCSRKPSDSSSGLAPNRNLATKSTINPRFAYLVTGTKGDGKRVKRLLKAIHHPRNYYL
LHLDLEASDERMELAKYVRSEKKKFENVMVMGLADLVTEKGPTMLASTLHGVAILLK
KAKDWDWFINLSASDYPLMPQDDILHIFSYLPRYLNFIEHTSNIGWKENQRARPIIID
PGFYHLKKSGVFWAKERRSLPASFKLFMGSTSVALTRPFLEFCIWGWDNLPRTLMYY
TNFLLSSEGYFQTVVCNNKDYQNTTVNHDLHYTKWDPLQQRTLNVTVENFRDMVQSGA
PFAREFREDDLVLDKIDIELLGQTDGTGLELKTPDVVKPTVSWKRLEKLMVRLLDHENF
RAKQCK"

repeat_region   complement(20872. .20971)
/rpt_family="(TGG)n"

```
repeat_region    K"
                 complement(33140..33196)
repeat_region    /rpt_family="(GAA)n"
                 33357..33398
repeat_region    /rpt_family="(GAA)n"
                 33469..33489
misc_feature     /rpt_family="AT_rich"
                 complement(34248..34296)
                 marginal_shadowexon"
                 35206..35258
misc_feature     /note="exon predicted by xgrail, quality
repeat_region    complement(35273..35541)
mRNA             /rpt_family="(GGA)n"
                 join(35801..35886,36023..36416,36517..37130,37242..37322)
gene             35801..37322
                 /gene="F13M22.12"
                 35801..37322
CDS              /note="identical to GB:U29480"
                 /gene="F13M22.12"
                 join(35827..35886,36023..36416,36517..37130,37242..37307)
                 /gene="F13M22.12"
                 /codon_start=1
                 /product="actin 3 protein"
                 /db_xref="PID:g3236244"
                 /translation="MADGEDIQPLVCDNGTGMVKAGFAGDDAPRAVPFSIVGRPRHTG
                 VMVGMQQKDAVVGDEAQSKRGILTLKYPIEHGIVNNWDDMEKIWHHTFYNELRVAPEE
                 HPILLTEAPLNPKANREKMTQIMFETFNAPAMYVAIQAVLSLYASGRTTGIVLDSGDG
                 VSHTVPIYEGYALPHAILRLDLAGRDLTDALMKILTERGYSFTTTAEREIVRDIKEKL
                 CYIALDYEQELETAKTSSSVEKNYELPDGQVITIGSERFRCPEVLYQPSMIGMENAGI
                 HETTYNSIMKCDVDIRKDLYGNIVLSGGTTMFPGIADRMSKEITALAPSSMKIKVVAP
                 PERKYSVWIGGSILASLSTFQQMWIAKAEYDESGPSIVHRKCF"
repeat_region    37700..37817
mRNA             /rpt_family="(CAAAA)n"
                 complement(<37864..>38967)
gene             /gene="F13M22.13"
                 complement(<37864..>38967)
CDS              /gene="F13M22.13"
                 complement(37864..38967)
                 /gene="F13M22.13"
                 /codon_start=1
                 /product="myb-related protein"
                 /db_xref="PID:g3236245"
                 /translation="MKERQRWGEEDALLRAYVRQFGPREWHLVSERMNKPLNRDAKS
                 CLERWKNYLKPGIKKGSLTEEEQRLVIRLQEKHGNKWKKIAAEVPGRTAKRLGKWWEV
                 FKEKQQREEKESNKRVEPIDESKYDRILESFAEKLVKERSNVVPAAAAAATVMANSN
                 GGFLHSEQQVQPNPVIPWLATSNNGNNVVARPPSVTLTLSPSTVAAAAPOPPIPWL
                 QQQOPERAENGPGGLVLGSMMPSCSGSSESVFLSELVECCRELEEGHRAWADHKKEAA
                 WRLRRLELQLESEKTCRQREKMEEIEAKMKALREEQKNAMEKIEGEYREQLVGLRRDA
                 EAKDQKLADQWTSRHIRLTKFLEQQMGCRLDRP"
repeat_region    complement(39531..39561)
repeat_region    /rpt_family="AT_rich"
                 complement(40737..40818)
repeat_region    /rpt_family="(TAA)n"
                 complement(41089..41112)
repeat_region    /rpt_family="POLY_A"
                 complement(41190..41213)
misc_feature     /rpt_family="POLY_A"
                 complement(41283..41310)
                 marginal_shadowexon"
repeat_region    complement(41558..41602)
repeat_region    /rpt_family="AT_rich"
                 complement(41966..42068)
repeat_region    /rpt_family="(GAAA)n"
                 complement(42281..42325)
misc_feature     /rpt_family="AT_rich"
                 42889..42968
misc_feature     /note="exon predicted by xgrail, quality excellent"
                 complement(43132..43154)
repeat_region    43371..43479
mRNA             /rpt_family="(TAAAA)n"
                 complement(join(43750..44452,45240..45558,45719..45963)))

gene             /gene="F13M22.14"
                 complement(43750..45963)
CDS              /gene="F13M22.14"
                 complement(join(44143..44452,45240..45558,45719..45878))
                 /gene="F13M22.14"
                 /codon_start=1
                 /product="putative expansin protein"
                 /db_xref="PID:g3236246"
                 /translation="MTATAFRVGLWLAVTASFLLTANAKIPGVSGGPWQNAHATFY
                 GGSDASGTMGGACGYGNLYSQGYGVNTAALSTALFNNGFSGCACFEIKCTDDPRWCVP
                 GNPSILVTATNFCPPNFAQPSDDGGWCNPPREHFDLAMPMFLKIGLYRAGIVPVSYKR
                 VPCRKIGGIRFTVNGFRYFNLVLTVNVAGAGDINGVSVKGSKTDWVRMSRNWGQNWQS
                 NAVLIGQSLSFRVTASDRRSSTSWNVAPATWQFGQTFSGKNFRV"
repeat_region    44795..44903
repeat_region    /rpt_family="AT_rich"
                 46287..46312
misc_feature     /rpt_family="AT_rich"
                 complement(47301..47444)
repeat_region    /note="exon predicted by xgrail, quality
                 excellent_shadowexon"
                 48559..48589
mRNA             /rpt_family="AT_rich"
                 <48689..>50845
gene             /gene="F13M22.15"
                 <48689..>50845
CDS              /gene="F13M22.15"
                 48689..50845
                 /gene="F13M22.15"
                 /codon_start=1
                 /product="SCARECROW-like protein"
                 /db_xref="PID:g3236247"
                 /translation="MITEPSLTGISGMVNRNRLSGLPDQPSHSFTPVTLYDGFNYNL
                 SSDHINTVVAAPENSVFIREEEEEDPADDFDRSDAVLGYISQMLNEEDMDDKVCMLQ
                 ESLDLEAAERSLYEAIGKKYPPSPERNLAFAERNSENLDRVVPGNYTGGDCIGFGNQG
                 IKPLSSGFTLDFRNPQSCSSILSVPQSNGLITIYGDGIDESSKNNRENHQSVWLFRRE
                 IEEANRFNPEENELIVNFREENCVSKARKNSSRDEICVEEERSSKLPAVFGEDILRSD
                 VVDKILVHVPGGSMKEFNALRDVLKKVEKKASDAQGGKRRARGRGRGRGRGGGGG
                 QNGKKEVVDLRSLLIHCAQAVAADDRRCAGQLLKQIRLHSTPFGDGNQRLAHCFANGL
                 EARLAGTGSQIYKGIVSKPRSAAAVLKAHQLFLACCPFRKLSYFITNKTIRDLVGNSQ
                 RVHVIDFGILYGFQWPTLIHRFSMYGSPKVRITGIEFPQPGFRPAQRVEETGQRLAAY
                 AKLFGVPPEYKAIAKKWDAIQLEDLDIDRDEITVVNCLYRAENLHDESVKVESCRDTV
                 LNLIGKINPDLFVFGIVNGAYNAPFFVTRFREALFHFSSIFDMLETIVPREDERMFL
                 EMEVFGREALNVIACEGWERVERPETYKQWHVRAMRSGLVQVPFDPSIMKTSLHKVHT
                 FYHKDFVIDQDNRWLLQGWKGRTVMALSVWKPESKA"
repeat_region    49591..49709
repeat_region    /rpt_family="(GAA)n"
                 49635..49689
misc_feature     /rpt_family="(GGA)n"
                 51112..51155
misc_feature     /note="exon predicted by xgrail, quality good"
                 complement(51301..51387)
                 marginal_shadowexon"
mRNA             complement(join(51428..51666,51782..51905,51988..52049,
                 52131..52264,52306..52342,52427..52698,52779..>53043))
gene             /gene="F13M22.16"
                 complement(51428..>53043)
CDS              /gene="F13M22.16"
                 complement(join(51547..51666,51782..51905,51988..52049,
                 52131..52264,52306..52342,52427..52698,52779..53043))
                 /gene="F13M22.16"
                 /note="unknown protein"
                 /codon_start=1
                 /db_xref="PID:g3236248"
                 /translation="MAMMTTTTTHFHPLLPANTYKSGAVASSFVSVPRSSSLQFRSL
                 VSDSTSICGPSKFTGKNRRVSVTVSAAATTEPLTVLVTGAGGRTGQIVYKLKERSEQ
                 FVARGLVRTKESEKINGEDEVFIGDIRDTASIAPAVEGIDALVILTSAVPQMKPGFD
                 PSKGGRPEFFFDDGAYPEQVDWIGQKNQIDAGVDDGLSFVVTAKAAGVKQIVLVGSM
                 GGTNINHPLNSIGNANILVWKRKAEQYLADSGIPYTIRAGGLQDKDGGIRELLVGKD
                 DELLETETRIARADVAEVCVQALQEEAKFKALDLASKPEGTGTPTKDFKALFTQVT
                 TKF"
mRNA             complement(join(<53450..54129,54196..54289,54371..54453,
                 54545..54596,54711..54868,54958..>56602))
```

```
gene            /gene="F13M22.17"
CDS             complement(<53450..>56602)
                /note="predicted by grail"
                /gene="F13M22.17"
CDS             complement(join(53450..54129,54196..54289,54371..54453,
                54545..54596,54711..54868,54958..56602))
                /gene="F13M22.17"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236249"
                /translation="MEDGNKLDRKKTMTMNWEGLGDFEDEDDDRFFETHDRLSSALTF
DMANASSDEDEDFDDCRLSFSSAVSSLLTTASRKFRTPAMSPDYDIWMAAPGSISERR
RRLLHGMGLASNKDMVSAVSIRRVVSNAPVVSNGEEKMKKKMMNHEVDEQDHDHPVM
LARSRSESDIERFLIEKRRKEEILGKISKQRLTRYSTICPTRTRICQYQTPIRQTPA
VCGNGKALRGGGEALTSVMSNARVGAFFLIKNLDTGKEFIVNEYDEDGMWNRLSDLQT
GKQLTLEEFEKCVGYSPVVKELMRRENVRNINYEPLMDLRKFNSYLSKSVRLSKRRGA
ALLKNIKGVAHSMSLRVADKDVSDGSTDSPKKGKDHKHGKANEWKVRPTGKSIYKELS
ALHMCQEIQAHEGAVWTIKFSQDAHYLASGGADRVIHWVEQVQECELMSMNEGSLTPIH
PSLCDSSGNEITVVEKKKKGKGSGSGRRHHNIPDYVHVPETVFSFSDKPVCSLKGHLDA
ILDLSWSKSQLLLSSSMDKTVRLMWDIETKTCLKLFAHNDVTCIQFSPVNENYFLSGS
LDAKIRIWSIQDRHVVEWSDLHEMVTAACYTPDQGALIGSHKGICRAYDTEDCKLSQ
TNQIVQSNKSQAKRKITSFQFSWNPSEVLVTSADSRIRILDGSEVIHKFKGFRNT
CSQLSASYSQDGKYIICASEDSQVYLWKNDFHRTRSTLTTQSHEHFCKDVSAAYPWH
GHVRGEPPVQIHSKRHSKRISTSSQPSSTISSPTKEETSATGPTTSNRNKKSGLPPM
PKKAATKSQIQPEEEAGPELGSSESFSRSSMNSSEQHSSRRFGESPSINTSSRLSSWSWF
DSGGHGPQTIQPTAWGMVIVTATVHGEIRSYQNFGLPRRIGRQTLF"
repeat_region   complement(56145..56194)
                /rpt_family="(GAA)n"
misc_feature    56686..56744
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    complement(56743..56808)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    complement(57092..57128)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    57429..57497
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    complement(57759..57834)
                /note="exon predicted by xgrail, quality excellent"
mRNA            complement(join(<57887..58388,59719..59846,60078..60132,
                60218..60291,60384..60527,61543..>61581))
                /gene="F13M22.18"
gene            complement(<57887..>61581)
                /gene="F13M22.18"
                /note="predicted by genefinder"
CDS             complement(join(57887..58388,59719..59846,60078..60132,
                60218..60291,60384..60527,61543..61581))
                /gene="F13M22.18"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236250"
                /translation="MPVRVVESNTPAQAFSGTQNVSNQQKEEAWRVNVQIGIDLEHG
YLCGTMEALNVPMADTPVITFWEGEIVDGKNYTFYTGKWEATREDDMRHWSKFPSFSP
LQRWKEQYFVNVGTDCGLIIAGFYVCFSCSDGISGFYDPNSSFHHMIISSSKNVL
KMEEVSVKKRKFQTDQSDELSLLPLSKHTCFANVACSENTNGNSEIDTEYSMSSYVN
STTSMECNNDIEMKEESSGSCGEDKMISFESHLDYIYGTQNLEDFSEKVIENILYLDE
QEEEEDAKGCSSNAAKFVLSSGRWTVNQGKENRILV"
repeat_region   58770..58795
                /rpt_family="(GA)n"
repeat_region   59165..59234
                /rpt_family="AT_rich"
repeat_region   59249..59312
                /rpt_family="AT_rich"
repeat_region   complement(59446..59478)
                /rpt_family="AT_rich"
misc_feature    complement(61713..61851)
                /note="exon predicted by xgrail, quality good_shadowexon"
mRNA            join(<62177..62406,62814..63035,63131..63242,63337..63399,
                63504..63642,63940..63995,64072..64145,64247..64394,
                64470..64652,64757..64844,64948..65058,65257..65354,
                65591..65708,65788..65882,66002..66108,66213..>66306)
                /gene="F13M22.19"
```

```
gene            <62177..>66306
                /gene="F13M22.19"
CDS             join(62177..62406,62814..63035,63131..63242,63337..63399,
                63504..63642,63940..63995,64072..64145,64247..64394,
                64470..64652,64757..64844,64948..65058,65257..65354,
                65591..65708,65788..65882,66002..66108,66213..>66306)
                /gene="F13M22.19"
                /codon_start=1
                /product="putative phosphoribosylaminoimidazole
                carboxylase"
                /db_xref="PID:g3236251"
                /translation="MLLLKQSSAAVLVVGNTPVLHTSRSTYRVGPFPVTRTQSFQSL
TMANLQKLPTSSSGKLNTASAVPCSSHDASPISENRENKHVHGVESKIVGLGGGQLG
RMLCQAASQLAIKVMLDPSKNCSASALSYGHMVDSFDDSATVEEFAKRCGVLTVEIE
HVDVDTLEKLEKQGVDCQPKASTIRIIQDKYMQKVHFSOHGIPLPEFMEISDIEGARK
AGELFGYPLMIKSKRLAYDGRGNAVANNQDELSSAVTALGFSRGLYIEKWAPFVKEL
AVIVARGRDGSMVCYPVVETHRDNICHIVAPADVPWKINKLATDVAQKAVGSLEGA
GVFAVELFLTEDSQILLNEVAPRPHNSGHQTIECCYTSQEFQHLRAVVGLPLGDPSMR
TPASIMYNILGEDDVIDGEAGFKLAHRLIARALCIPGASVHWYDKPEMRKQRKMGHIT
LVGQSMGLIEGYLQCLLSEQHQVHETPRVAIIMGSDTDLPVMKDAAKILDLFGVTHE
VKIVSAHRTPEMMTYATSAHSRGVQVIIAGAGAAHLPGMVASLTPLPVIGVPVRAT
RLDGVDSLLSIVQMPRGVPVATVAINNATNAALLAVRMLGISDTDLVSRMRQYQEPMR
DENLNKGEKLETEGWESYLNQ"
mRNA            complement(join(<66647..66829,67056..67229,67307..67405,
67776..67976,68080..68187,68273..68651,68917..69175,
69620..70064,70183..>70242))
                /gene="F13M22.20"
gene            complement(<66647..>70242)
                /gene="F13M22.20"
CDS             complement(join(66647..66829,67056..67229,67307..67405,
67776..67976,68080..68187,68273..68651,68917..69175,
69620..70064,70183..70242))
                /gene="F13M22.20"
                /codon_start=1
                /product="CER1-like protein"
                /db_xref="PID:g3236252"
                /translation="MASRPGFLTDWPWTPLGSFKILLAPLVFDSIYSYATIRDHEKL
LIVAVTWRIVHSQIWISLSRYQTAKGTKRILNKSIEFDQVDRERTWDDQIIFNTLIY
YLTKVYVSGTSTIPFWRTDGVILVALLHAGPVEFIYWFHRALHHHFLYSRYHSHHHS
SIVTEPITLCATNSKFPWVLIVAVVHPFAEHIGYTLILGLPLITTFMCGTVSVSIALY
LTYIDFMNNMGHCNFELIPKFLSLLPLKFCLYTDFLSFSFHSHHTQFRTNYSLFMMYD
YIYGTTDECSDSLVETSLEKEEKPDAIHLTHLTSLDSIYHLRLGFASLSSHPLSSRC
YLFLMKPFALILSFILRSFSFQTFVVERNRFRDLITLHSHLLPKFSSHYMSHQQKECIN
KMIEAAILEADKKGVKVMSLGLLNQGEELNYGEMYVRRHPKLKIRIVDGGSLAAEVV
LHSIPVGTKEVLFRGQITKVARAIVFSLCQNAIKVMVLRKEEHSMLAEFLDDKCKENL
VLTNNYPMIWLVGDGCLSTKEQKMAKDGTFLPFSQFPPKTLRKDCFYHTTPAMIIPH
SAQNIDSCENWLGRRVMSAWRVGGIVHALEGWKEHECGLDDNSIINPPRWEAALRNG
FQPLLLPSLET"
repeat_region   complement(67680..67715)
                /rpt_family="AT_rich"
misc_feature    complement(70565..70634)
                /note="exon predicted by xgrail, quality marginal"
mRNA            complement(70808..73053)
                /gene="F13M22.21"
gene            complement(70808..73053)
                /note="highly similar to GB:X95909(polymorphism exists
                between nucleotides 73005 and 72085 which is a GA repeat.
                We found 6 copies in our sequence whereas only 5 copies
                exist in GB:X95909)"
                /gene="F13M22.21"
CDS             complement(71000..73027)
                /gene="F13M22.21"
                /codon_start=1
                /product="receptor-like protein kinase"
                /db_xref="PID:g3236253"
                /translation="MFLKLLTIFFFFFNLIFQSSSSQSLNFAYNNGFNPPTDLSIQGI
TTVTPNGLLKLINTVQKTGHAFYTKPIRFKDSPNGTVSSFSTSFVFAIHSQIAILSG
HGIAFVVAPNASLPYGNPSQYIGLFNLANNGNETNHVFAVELDTIILSTEFNDTNDNHV
GIDINSLKSVQSSPAGYWDEKGQFKNLTLISRKPMQVWVDYDGRTNKIDVTMAPFNED
KPTRPLVTAVRDLSSVLLQDMYVGFSSATGSVLSEHYILGWSFGLNEKAPPLALSRLP
KLPRFKRISEFYIKGMPLLSLFLFFSFIFLVCYIVRRRRKFAEELEEWEXEFGKNR
FRFKDLYYATKGFKEKGLLGTGGFSGSVYKGVMPGTKLEIAVKRVSHESRQGMKEFVAE
IVSIGRMSHRNLVPLLGYCRRRGELLLVDYMPNGSLDKYLNTPEVTLNWKQRIKVI
```

```
                LGVASGLFYLHEEWEQVVIHRDVKASNVLLDGELNGRLGDFGLARLYDHGSDPQTTHV
                VGTLGYLAPEHTRTGRAMATDVFAFGAFLLEVACGRRPIEFQQETDETFLVDWVFG
                LWNKGDILAAKDPNMGSECDEKEVEMVLKLGLLCSHSDPRARPSMRQVLHYLRGDAKL
                PELSPLDLSGSGMFGVHDGFSELGMSYSSSSVFGFTGGSSIADSQLSSGR"
repeat_region   72962..73005
                /rpt_family="(GAA)n"
repeat_region   complement(73347..73473)
                /rpt_family="AT_rich"
mRNA            join(<74148..74206,74511..74566,75952..>77285)
                /gene="F13M22.22"
gene            <74148..>77285
                /gene="F13M22.22"
CDS             join(74148..74206,74511..74566,75952..77285)
                /gene="F13M22.22"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236254"
                /translation="MVQIVPHVANAHMVKTSRKGLLALVVVGVEECCPTRRCDLRDPN
                LVVDSNCLLSRSTHQSSLVEDSIIINHFRILNFNNRASIFLCLRIGVICSTQSKMYGG
                KSYILKGSVSLSLILILVTTLLVSAKNQPRISLIEVKTCNLAKGEWVEDKKRPLYS
                GFECKQWLSNIFSCRVMGRPDFSFEGYRWQPEGCNIPEFNRVNFLRRMQNKTIAFIGD
                SLGREQFQSLMCMATGGKSPSPEVGVSEYGLVIPKGAPRPGGWAYRFPTTNTTVLSY
                WSASLTDLVPMNNTDPPHLIAMHLDRPPAFIRNYLHRFHVLVLNTGHHWSRDKIEKNH
                WVMHVNGTRVEGGFKNVENAKIFTIHSLVKWLDAQLPHPRLKAFFTTISPRHEKCN
                NTIPLSRGSKITGEGGSLDTIVESAVNGTRVKILDITALSKLRDEAHIAGCKLKPKKA
                SNVTSAPTFNDCLHWCLPGIPDTWNELLIAQL"
repeat_region   complement(75409..75618)
                /rpt_family="AT_rich"
repeat_region   complement(77388..77424)
                /rpt_family="AT_rich"
repeat_region   complement(77464..77524)
                /rpt_family="AT_rich"
misc_feature    77550..77635
                /note="exon predicted by xgrail, quality
                excellent shadowexon"
mRNA            complement(join(<78252..78416,78604..79152,79555..79945,
                80111..>80604))
                /gene="F13M22.23"
gene            complement(<78252..>80604)
                /note="predicted by genscan"
                /gene="F13M22.23"
CDS             complement(join(78252..78416,78604..79152,79555..79945,
                80111..80604))
                /gene="F13M22.23"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236255"
                /translation="MAIFDPFKLHHKPPSIFPATTVKPSHVLSLVAKFFTICIFISV
                AMISFYIIFFGCSDCHHITGLRRLGPENVNLTTNSSASSSTIRRPNQSSEATDISHIA
                FGIGGSIQTWRDRSRYSELWWRPNVTRGFIWLDEEPPLNMTWLSTSPPYQVSADTSRF
                SYTCWYGSRSAIRMARIIKETFELGLTDVRWFIMGDDDTVFFVDNLITVLNKYDHNQM
                EACLSEIGVPLTKELGFHQVDIRGNPYGLLAAHPVAPLVTLHHLDYVDPIFPGTTQID
                ALRRLVSAYKTDPSRIIQHSFCHDQTRNWVVSVSWGYTIQIYPLVTAKELETPFLTF
                KSWRTSSEPFSPFTDRPISEDPCERPLVYFLDRVYEVGSSQTLITYRKHVEVGESTQC
                NSPDYSRANPVEFIDVSTTLTPDLWKMAPRRQCCEIVNSEEDSESVINVKIRHFNPL
                ESVTPQSKHPNDEWTDSVLSTETI"
repeat_region   complement(79324..79361)
                /rpt_family="AT_rich"
repeat_region   complement(80698..80727)
                /rpt_family="AT_rich"
misc_feature    complement(81151..81260)
                /note="exon predicted by xgrail, quality excellent"
misc_feature    82037..82089
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
misc_feature    82508..82558
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
mRNA            complement(<83772..>84686)
                /gene="F13M22.24"
gene            complement(<83772..>84686)
                /gene="F13M22.24"
```

```
CDS             complement(83772..84686)
                /gene="F13M22.24"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236256"
                /translation="MEKPGGFWIPKKSNKESSWEELAFAEDDAAGSLWPPSRYTCSFC
                RREFKSAQALGGHMNVHRDRARLKQADDQYLFPKSSSSSPEYPSHKDSDNIHETSCYT
                LVFNTKPNYFKTQHSCVIDLSSSSLPYLTPSRVSSGLPGKQHTSSSSPSFVVEPSKN
                SKYIPSSSPWSCPSTVVEQKSCDIYEIPAMEGEKKRKTESDVPRIGHKAKLSLGNTTD
                LSVSMNLVIHQSFPITAHGSDEEIGRVDIRKRRRHESPSQQSIFISSLSCKSDIITR
                NEESKHKGDRFEDLDLELRLGTDPPKGI"
                84230..84321
                /rpt_family="(GAA)n"
repeat_region   84910..84962
                /rpt_family="AT_rich"
repeat_region   85196..85220
                /rpt_family="(TA)n"
repeat_region   complement(85246..85307)
                /rpt_family="AT_rich"
misc_feature    complement(85742..85810)
                /note="exon predicted by xgrail, quality marginal"
repeat_region   86102..86122
                /rpt_family="AT_rich"
misc_feature    86195..86308
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   86660..86687
                /rpt_family="AT_rich"
misc_feature    86972..87002
                /note="exon predicted by xgrail, quality marginal"
mRNA            87049..87569
                /gene="F13M22.25"
gene            87049..87569
                /gene="F13M22.25"
CDS             87120..87332
                /gene="F13M22.25"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236257"
                /translation="MQIRYGNESQQLSSSYYVRLPGDSGRFVRSDLDVSGPGSNSSNR
                ASSSTSTTTFAVYAGVQSPVSRQVVT"
mRNA            join(<88062..88169,88258..88308,88608..88684,88770..88857,
                88934..89209,89301..>89585)
                /gene="F13M22.26"
gene            <88062..>89585
                /gene="F13M22.26"
CDS             join(88062..88169,88258..88308,88608..88684,88770..88857,
                88934..89209,89301..89585)
                /gene="F13M22.26"
                /codon_start=1
                /product="putative alcohol dehydrogenase"
                /db_xref="PID:g3236258"
                /translation="MAAPIRFFELNTGAKLPCVGLGTYAMVATAIEQAIKIGYRHIDC
                ASIYGNEKEIGGVLKKLIGDGFVKREELFITSKLWSNDHLPEDVPKALEKTLQDLQID
                YVDLYLIHWPASLKKESLMPTPEMLTKPDITSTWKAMEALYDSGKARAIGVSNFSSKK
                LTDLLNVARVTPAVNQVECHPVWQQGLHELCKSKGVHLSGYSPLGSQSGKGEVRLKVL
                QNPIVTEVAEKLGKTAQVALRWGLQTGHSVLPKSSSGARLKENLDVFDWSIPEDLFT
                KFSNIPQASSFIFLFFSR"
                89723..89798
                /note="exon predicted by xgrail, quality excellent"
misc_feature    complement(89743..89842)
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
repeat_region   complement(90572..90624)
                /rpt_family="AT_rich"
mRNA            join(<90955..91074,91214..91264,91485..91561,91649..91736,
                91942..92217,92291..>92551)
                /gene="F13M22.27"
gene            <90955..>92551
                /gene="F13M22.27"
CDS             join(90955..91074,91214..91264,91485..91561,91649..91736,
                91942..92217,92291..92551)
                /gene="F13M22.27"
                /codon_start=1
```

```
                    /product="putative alcohol dehydrogenase"
                    /db_xref="PID:g3236259"
                    /translation="MANAITFFKLNTGAKFPSVGLGTWQASPGLVGDAVAAAVKIGYR
                    HIDCAQIYGNEKEIGAVLKKLFEDRVVKREDLFITSKLWCTDHDPQDVPEALNRTLKD
                    LQIEYVDLYLIHWPARIKKGSVGIKPENLLPVDIPSTWKAMEALYDSGKARAIGVSNF
                    STKKLADLLELARVPPAVNQVECHPSWRQTKLQEFCKSKGVHLSAYSPLGSPGTTWLK
                    SDVLKNPILNMVAEKLGKSPAQVALRWGLQMGHSVLPKSTNEGRIKENFNVFDWSIPD
                    YMFAKFAEIEQVRS"
  repeat_region     complement(91371..91445)
                    /rpt_family="(CAAAA)n"
  misc_feature      92640..92726
                    /note="exon predicted by xgrail, quality excellent"
  misc_feature      complement(93519..93682)
                    /note="exon predicted by xgrail, quality good_shadowexon"
  misc_feature      93640..93819
                    /note="exon predicted by xgrail, quality excellent"
  repeat_region     complement(93688..93779)
                    /rpt_family="(CAT)n"
BASE COUNT   29945 a   17475 c   16793 g   29632 t
ORIGIN

Query Match                    7.1%;   Score 70.4;   DB 39;   Length 93845;
Best Local Similarity         74.2%;   Pred. No. 4.9e-05;
Matches   89;   Conservative   0;   Mismatches   31;   Indels   0;   Gaps   0;

Qy    363  atctctggcctccaagatcttatacatgtagcttttgtaaaagggaatttagatctgc  422
            |||| ||  |||| ||||| || ||||||  || || ||  |||| | |||| |||||
Db  84594  ATCTTTGGCCGCCAAGATCTTATACGTGTAGCTTCTGCCGAGAGAATTTAAATCCGGC  84535

Qy    423  tcaagctcttgggacacatgaatgttcatagaagagatagagccatttggagacaatc  482
            | |||| ||  || ||||| ||||||||| |||||  ||| || |  ||| |||| |
Db  84534  TCAAGCCCTCGGTGGCCATATGAACGTTCACCGAAGACAACAGCACGCCTTAAGCAAGC  84475


RESULT  6
DDU32174
LOCUS       DDU32174    5931 bp    DNA         INV       22-JAN-1997
DEFINITION  Dictyostelium discoideum spore lysis A (splA) gene, complete cds.
ACCESSION   U32174
NID         g974333
KEYWORDS    .
SOURCE      slime mold.
  ORGANISM  Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyostelida;
            Dictyostelium.
REFERENCE   1  (sites)
  AUTHORS   Tan,J.L. and Spudich,J.A.
  TITLE     Developmentally regulated protein-tyrosine kinase genes in
            Dictyostelium discoideum
  JOURNAL   Mol. Cell. Biol. 10 (7), 3578-3583 (1990)
  MEDLINE   90287147
REFERENCE   2  (bases 1 to 5931)
  AUTHORS   Nuckolls,G.H., Osherov,N., Loomis,W.F. and Spudich,J.A.
  TITLE     The Dictyostelium dual-specificity kinase splA is essential for
            spore differentiation
  JOURNAL   Development 122 (10), 3295-3305 (1996)
  MEDLINE   97053827
REFERENCE   3  (bases 1 to 5931)
  AUTHORS   Nuckolls,G.H. and Spudich,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1995) Glen H. Nuckolls, Departments of
            Biochemistry and Developmental Biology, Stanford University, B400
            Beckman Center, Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1..5931
                     /organism="Dictyostelium discoideum"
                     /strain="JH10"
                     /db_xref="taxon:44689"
                     /chromosome="4"
     gene            861..5763
                     /gene="splA"
     CDS             join(861..1001,1083..1156,1224..5763)
```

```
                     /gene="splA"
                     /standard_name="spore lysis A"
                     /note="exhibits tyrosine and serine/threonine
                     autophosphorylation activity; also called DPYK1"
                     /codon_start=1
                     /product="non-receptor tyrosine kinase"
                     /db_xref="PID:g974334"
                     /translation="MSFKFFKKLVPSNEYRWDLRKSNSLTLNIEDKSRCSYRLPTSG
                     SKGIAKSTQPFSSSFYFELFITTNGNGDKICFGLTTNDHPIEVYPGNYQGSYGYSGDG
                     KCYFGTNEGRVYGPSFSSGDVVGCGYDSSSKTLYFTKNGVYLGVAAQKVNLIGLYPTV
                     GLQNPGEGVVINFFGPFSYRGAPEKPSKQSTIKDSGGSSIIPSEDLIPKEEFVCRWS
                     EKKNYHGKHVVVNRTAFLPLDSPKDTIGGVRATQPFGEGFCYFEVIIDQLKDGQLSI
                     GLANLEYPTFYHVGWMPRSYGYHNDDGRKFRWREEPGVNEGESYGSSYKKGDIIGCGL
                     SFTSREIFFTKNGMYLGTAFSNVYGVFYPSVAFNEPGISITGVPGPPFKFSQVTLMLK
                     NVNSTSLVPNGNNNNSNNNNNNNNIGNGKITTTTTSPSSINNNEDISSNN
                     SNSNSNSNNNINNNNNNNNNNNIYLTKKPSIGSTDESSTGSLGGNNSSGNNNSSSG
                     SIGNNSSIIKQRSPPHSINGPLMLPPSSTNNNNIYSSYNSTTAGSSTTILPTLNHPI
                     FGNTTSNNNSSSTLSVGGNNLLGRHCQSLPITASTNHTLSSLGVSFSSPSSSPKTS
                     PRKIVNSSEDLGVQTFQDQQPPSAWRRCGKSIIKTKDDITLTIITIIKKKTSVAMADRP
                     FSSNSSSTICYFEVYLEGHDKKGSITVGLSHSTYPFIKHIGREPKSYGFSSEGEKYGG
                     SEIGEPYGPFFFFDGDSIASSCVIGCGINTSTRDIFFTKNGHYLGVAFSRVTSDPLYP
                     SISFRGVVGGLCVATFPGGHFRFNIEDLPGISPSVWTEALGPDRQGSGFKNWAPNDVA
                     IWLESFNYGQYRKNFRDNNISGRHLEGITHAMLKNDLGIEPYGHREDIINRLNRMIQI
                     WNDKSPDSYPKIAIDSSDKIRWPASGGSSGINISGGVVIGSGSGDGITPEISSSSK
                     NIRPYKSYTQKEIEDRNRRSTISGGEKKNKYIDNQMDHQISMDSDGLLPFGCGGF
                     PDEKNSSKTLSNEQIRYLQQRKDEPPIAISSTGNGGSVSSTGGSSGFLTFPSSNSLTH
                     PPQRDKPTQEFTHLPPITSNYKGITNTGQPHKSFDQPLELFPRHSAFSNNGNNGNNN
                     NNNNNIKANQQQQSYQQSOTQQQQHITSTSTSTNKWIDPFGGWETQSSLSHP
                     PSRPPPPPPPPQLPVRSEYEIDFNELEFGQTIGKGFFGEVKRGYWRETDVAIKIIYR
                     DQFKTKSSLVMFQNEVGILSKLRHPNVVQFLGACTAGGEDHCIVTEWMGGGSLRQFL
                     TDHFNLLEQNPHIRLKLALDIAKGMNYLHGWTPPILHRDLSSRNILLDHNIDPKNVY
                     SSRQDIKCKISDFGLSRLKKEQASQMTQSVCIPYMAPEVFKGDNSEKSDVYSYGMV
                     LFELLTSDEPQQDMKPMKMAHLAAYESYRPPILPTTSSKWKEILTQCWDSNPDSRPTF
                     KQIIVHLKEMEDQGVSSFASVPQTIDTGVYA"
     intron          1002..1082
                     /gene="splA"
                     /note="intron confirmed by sequencing RT-PCR product"
     intron          1157..1223
                     /gene="splA"
                     /note="intron confirmed by sequencing RT-PCR product"
     conflict        5259
                     /gene="splA"
                     /note="reported as C in mRNA sequence, GenBank Accession
                     Number M33785"
                     /citation=[1]
                     /replace="c"
     conflict        5310..5311
                     /gene="splA"
                     /note="reported as GT in mRNA sequence, GenBank Accession
                     Number M33785"
                     /citation=[1]
                     /replace="gt"
BASE COUNT   2247 a   857 g   1790 t
ORIGIN

Query Match                    4.9%;   Score 49;   DB 3;   Length 5931;
Best Local Similarity         47.3%;   Pred. No. 1.1;
Matches  148;   Conservative   0;   Mismatches  165;   Indels   0;   Gaps   0;

Qy    216  caacagtaagtacttcaaaacaacaaagcatcatggcaacaacagatggagtacttgaataa  275
            ||||| ||  ||||| || || |||||||||||| ||| ||||| || || || || ||
Db    722  CAACAACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA  781

Qy    276  caacaatggcgcaataacaacaacaataatgttacaagctcattacgagataattatgg  335
            |||| || | ||| ||| || || ||| || || || |||||| || || ||| ||||
Db    782  CAATAATAATAACAACAATAACAACAACAACAACAACAACCACCACCAACAACAAAAGAGG  841

Qy    336  aaatgaagatcatttacttggtggactattccttggcctccaagatcttatacatgtag  395
            |||| ||||||||  || || || ||| ||||| || ||||| |||||| |||||||||
Db    842  AGATTCTAGTTGGTTTAATATGAGCTTTAAATTCTTTAAAAGAAGAAATTGGTCCCATCCAA  901
```

```
Qy    396  ctttgtgaaagggaattagatctgctcaagctcttggtggacacatgaatgttcatag  455
           |  |||||||||| ||| || | ||||  || | | |||||||| || |||| |||
Db    902  TGAATATAGATGGGATTTAAGAAAATCAAATTCACTTACATTAAATATTGAAGATAAAAG  961

Qy    456  aaggatagaggccatttgagacaatcaccaccctagagatattaataggtattcctcct  515
           |||    | | |  || ||| | | | ||| || |  |||  ||| |||||||||  ||
Db    962  TAGATGTTCTTACAGACTACCAACATCTGGTTCAAAAGGTGTAAGTATTTTTTCTTTTTT  1021

Qy    516  aaaccttaatctt  528
           ||| ||| || ||
Db   1022  AAAATTTTATTT  1034
```

RESULT   7
AE000787/c
LOCUS   AE000787   38829 bp   DNA   BCT   10-DEC-1997
DEFINITION Borrelia burgdorferi plasmid lp38, complete plasmid sequence.
ACCESSION   AE000787
NID   g2690175
KEYWORDS   .
SOURCE   Lyme disease spirochete.
  ORGANISM Borrelia burgdorferi
            Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
            burgdorferi group.
REFERENCE   1 (bases 1 to 38829)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
            Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
            Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
            Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
            Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
            Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
            Artlach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
            Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
  TITLE   Genomic sequence of a Lyme disease spirochaete, Borrelia
            burgdorferi
  JOURNAL   Nature 390 (6660), 580-586 (1997)
  MEDLINE   98065943
REFERENCE   2 (bases 1 to 38829)
  AUTHORS   Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
            Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
            Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
            Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
            Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
            Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
            Artlach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
            Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-DEC-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES   Location/Qualifiers
     source   1..38829
                     /organism="Borrelia burgdorferi"
                     /db_xref="taxon:139"
                     /plasmid="lp38"
     gene   482..667
                     /gene="BBJ01"
     CDS   482..667
                     /gene="BBJ01"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ01"
                     /db_xref="PID:g2690206"
                     /translation="MKYHIVSIFVFLFLNACNPDSNTNQNNSKKELKTGRIPNKQIK
                     NALLDDLKNLIETASAG"
     gene   927..1211
                     /gene="BBJ02"
     CDS   927..1211
                     /gene="BBJ02"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ02"
                     /db_xref="PID:g2690207"
                     /translation="MINHLLCLKKETLDTLTVSDLKLKNSFEKLLSIKTDISKMLEQ
                     LLDVKDYINSTKTDIAKLESHLTELYKQIVKKSFQTTELKNNILSISNL"
     gene   1593..1745
                     /gene="BBJ03"
     CDS   1593..1745
                     /gene="BBJ03"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ03"
                     /db_xref="PID:g2690184"
                     /translation="MKAKKHPYYYNFLNVSLTLGIGSFFWPRGLYWQHGELLRSQVL
                     GEIYLR"
     gene   complement(2268..2381)
                     /gene="BBJ04"
     CDS   complement(2268..2381)
                     /gene="BBJ04"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ04"
                     /db_xref="PID:g2690185"
                     /translation="MLLYKKTFSMIGFCFRALSENEKHFLMLLFSAGTIF"
     gene   complement(2765..3261)
                     /gene="BBJ05"
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact.; similar to
                     GP:2228714 percent identity: 94.44; identified by sequence
                     similarity; putative;transposase-like protein, authentic
                     frameshift"
     gene   3486..3632
                     /gene="BBJ06"
     CDS   3486..3632
                     /gene="BBJ06"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ06"
                     /db_xref="PID:g2690186"
                     /translation="MISSKSFIIFLLSYLENYCVLFPFLIYLKNYMQSLSQAGHYK
                     ELNY"
     gene   complement(4164..4307)
                     /gene="BBJ07"
     CDS   complement(4164..4307)
                     /gene="BBJ07"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ07"
                     /db_xref="PID:g2690187"
                     /translation="MEMNLLINRFIFILNIFFDDINRGVPLLYRDKAIASRPKIRFSA
                     ILI"
     gene   4576..5496
                     /gene="BBJ08"
     CDS   4576..5496
                     /gene="BBJ08"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ08"
                     /db_xref="PID:g2690188"
                     /translation="MFLYTLTIGLISCNLDSKLPNKEQKNNDIKETLGSSVQENAL
                     NNLYGNQEEKDKFKNFEELKDESLIALAKSLASTRPTVGNIESAVLPVGHVVSLETS
                     ANKVSIPTISIKHNQKKEIKKEDLSPSTKEEKKADKAIKDIENLIRDSGFPELIESMY
                     SLKHEYTLIRNNFYDVITKIRNKKTSLIKNGRNNRDKIKELTQLQNNLKIVDELDEIM"
```

```
                     VHIDIAEQEIRSAAFFNEAKELLKEGIIKRLESENKVASQLARQALNKVEDALKSLE
                     ASSSKRGLAMGRRIIKELIENAKTVLSKS"
gene             6089..6862
                 /gene="BBJ09"
CDS              6089..6862
                 /gene="BBJ09"
                 /note="similar to percent identity: 99.61; identified by
                 sequence similarity; putative"
                 /codon_start=1
                 /transl_table=11
                 /product="outer surface protein D (ospD)"
                 /db_xref="PID:g2690176"
                 /translation="MKKLIKILLSLFLLLSISCVHDKQELSSKSNLNNQKGYLDNEG
                 ANSNYESKKQSILSELNQLLKQTNSLKEAKNTTDNLNASNEANKVVEAVINAVNLIS
                 SAADQVKSATKNMHDLAQMAEIDLEKIKNSSDKAIFASNLAKEAYSLTKAAEQNMQKL
                 YKEQQKISESESESDYSDSAEIKQAKEAVEIAWKATVEAKDKLIDVENTVKETLDKIK
                 TETTNNTKLADIKEAAELVLQIAKNAKEIVQEVVALLNT"
gene             7270..7476
                 /gene="BBJ10"
CDS              7270..7476
                 /gene="BBJ10"
                 /note="hypothetical protein; identified by Glimmer;
                 putative"
                 /codon_start=1
                 /transl_table=11
                 /product="B. burgdorferi predicted coding region BBJ10"
                 /db_xref="PID:g2690189"
                 /translation="MGARDEFKNGIRKLDEKIERARIELSTKTDVRSELSKQVKGLI
                 FPFYWILGIFIPAVIGLFLYLLQK"
gene             complement(7780..7965)
                 /gene="BBJ11"
CDS              complement(7780..7965)
                 /gene="BBJ11"
                 /note="hypothetical protein; identified by Glimmer;
                 putative"
                 /codon_start=1
                 /transl_table=11
                 /product="B. burgdorferi predicted coding region BBJ11"
                 /db_xref="PID:g2690190"
                 /translation="MFLSFCNLCITKNIHNIGSLRRVYTNFVYDYAFWLSKIAEDYFP
                 DAKDIVCIRGIKKKPYI"
gene             complement(8633..8725)
                 /gene="BBJ12"
CDS              complement(8633..8725)
                 /gene="BBJ12"
                 /note="hypothetical protein; identified by Glimmer;
                 putative"
                 /codon_start=1
                 /transl_table=11
                 /product="B. burgdorferi predicted coding region BBJ12"
                 /db_xref="PID:g2690191"
                 /translation="MVLGIWKKAENYVKFDELIIVKNKELFVI"
gene             complement(8877..9155)
                 /gene="BBJ13"
CDS              complement(8877..9155)
                 /gene="BBJ13"
                 /note="hypothetical protein; identified by Glimmer;
                 putative"
                 /codon_start=1
                 /transl_table=11
                 /product="B. burgdorferi predicted coding region BBJ13"
                 /db_xref="PID:g2690192"
                 /translation="MLHINHANEAFKEISSAKDKIDSSDIGKGIVDVQKTHNRLHDLL
                 RDIKTEFYSQKNSFLNGVRAEKSKKKNQILFRSTSSSINGKIRNANYIL"
gene             9125..9286
                 /gene="BBJ14"
CDS              9125..9286
                 /gene="BBJ14"
                 /note="similar to PID:62180 percent identity: 33.33;
                 identified by sequence similarity; putative"
                 /codon_start=1
                 /transl_table=11
                 /product="conserved hypothetical protein"


                 /db_xref="PID:g2690177"
                 /translation="MLHLHDLYATNFVIFQLSIYNLFVTFFKFSSNAEPDPLLVPFDL
                 LLFKDSSKP"
gene             complement(10040..10168)
                 /gene="BBJ15"
CDS              complement(10040..10168)
                 /gene="BBJ15"
                 /note="hypothetical protein; identified by Glimmer;
                 putative"
                 /codon_start=1
                 /transl_table=11
                 /product="B. burgdorferi predicted coding region BBJ15"
                 /db_xref="PID:g2690193"
                 /translation="MSISWLFIFKFISQTIDNVQKININKTGVFVALIKLKLNI"
gene             complement(10518..11105)
                 /gene="BBJ16"
CDS              complement(10518..11105)
                 /gene="BBJ16"
                 /note="similar to GB:U03641 PID:458218 percent identity:
                 35.19; identified by sequence similarity; putative"
                 /codon_start=1
                 /transl_table=11
                 /product="conserved hypothetical protein"
                 /db_xref="PID:g2690178"
                 /translation="MKIKAEKDKEALFSNRFGDCNEETNLNDDQDKELANYNNLKFGI
                 KYNLKDDINNKIQRMKILYEIKQELYDFGFARFNDFIKSFEVAKSQAYRYLKIYQK
                 VLEGKVSIDKIKEVGFKAILRDIKAKDSLNEDNHSESEGANESIPIRILVKDKELYNF
                 CKQDTKRLYFIIEKIYKEKRDVLSELIIEYEKNK"
gene             complement(11152..11889)
                 /gene="BBJ17"
CDS              complement(11152..11889)
                 /gene="BBJ17"
                 /note="similar to GP:2182756 percent identity: 39.43;
                 identified by sequence similarity; putative"
                 /codon_start=1
                 /transl_table=11
                 /product="conserved hypothetical protein"
                 /db_xref="PID:g2690179"
                 /translation="MDRKKSNITTIANLKGGVGKSTLSILFSVLKDLGKKVLLIDMD
                 SQNALTSYFRKYVFNFDKNNIYNLLIGNVYFDQCISKINDNIFIIPSHFLDEFNDKN
                 LDNKENLLSFCLDKNVLGHDFDYIFDTPPSFSFILKNALNTTNHIIIPVQPETWSIE
                 SLEILMKKITDKSYNISIVVNQFIKNRNILKEVEDALYKRYSNYIKGKIHYYNSIKVF
                 IINRLEPDIKSKYYKEAKDVLKNILDL"
gene             complement(11862..12452)
                 /gene="BBJ18"
CDS              complement(11862..12452)
                 /gene="BBJ18"
                 /note="similar to GB:U03641 PID:458217 percent identity:
                 39.38; identified by sequence similarity; putative"
                 /codon_start=1
                 /transl_table=11
                 /product="conserved hypothetical protein"
                 /db_xref="PID:g2690180"
                 /translation="MNDLLEKLKARKKELISKKEAEYNNNINKGTKERTAFFRIEEID
                 NKKIYYTKIFKYLVKFRIANKDNKLSLIFQKLNNKKNYLFNLFPLKEDNKFLGIKYG
                 WDKLEKPFFLRQNNKSYVKKLYLEFKFSKGSIKCYVQSLRTLLRKKDKESTKYYKF
                 NLEHIKRMENTVKFYSKKLKNKGVIKWIEKNQIL"
gene             complement(12445..13440)
                 /gene="BBJ19"
CDS              complement(12445..13440)
                 /gene="BBJ19"
                 /note="similar to GP:1655797 percent identity: 27.04;
                 identified by sequence similarity; putative"
                 /codon_start=1
                 /transl_table=11
                 /product="conserved hypothetical protein"
                 /db_xref="PID:g2690181"
                 /translation="MALNLNLNHNHNKTNTNKLLSDSRLKKIISVINYLNKSFEKKYD
                 ISIYRIHFNSEKLKELYPHHQIDILRLVLNSNISKEGKPTVIRTLREDLRFLIHIKAI
                 EKILTFSNNLGKFKGKLCIYKVSPIAVKLISAYFEAYKADLIRKIKKSKDGLDTQNV
                 TKNVTVIYNYHKNIYNKNSIETVFKKSYTKKKTKKNKKEFTKNNLEKRLKLPEEITK
                 EIISIAKKTKNPDKTYKNTLFNVKDFLNYLSYDYKKEDISYFFLSKLKEYKNKIHFMR
                 KYAPYKTDFYLLAGEFKDSYHSKWKTNKKTNFSGHVKEIANNILSKIIEKELKFE"
```

```
gene    complement(13772..13936)
        /gene="BBJ20"
CDS     complement(13772..13936)
        /gene="BBJ20"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ20"
        /db_xref="PID:g2690208"
        /translation="MQKAKHTDRTNLENLLNSFNKMNKNSNIHVQNKTTRNRKEIGSP
        NYIKTIKVY"
gene    15514..15660
        /gene="BBJ21"
CDS     15514..15660
        /gene="BBJ21"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ21"
        /db_xref="PID:g2690209"
        /translation="MARAIKYLHISFKYVLLKPILVLIKPRKLSPMFLFILKKKQKAC
        LYIF"
gene    complement(15902..16003)
        /gene="BBJ22"
CDS     complement(15902..16003)
        /gene="BBJ22"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ22"
        /db_xref="PID:g2690211"
        /translation="MELLKMKKVEYVLYSKDKKHTILLDANILQGKS"
gene    16505..17329
        /gene="BBJ23"
CDS     16505..17329
        /gene="BBJ23"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ23"
        /db_xref="PID:g2690211"
        /translation="MNGVIMREISCCFLLLTFSVVCVYSFDVSSRKFYGILEGYYSGK
        IEELSKKNDEDVYIYRFGKRFKETLSEMSSGIKSYFFNLVDYQIARLLQNKEGRRNSSK
        SYSVLKSTQKSLLELIASTDFKGLEKTIQSDICRILGDVNLMLLFRYAGGAALSKLANE
        ARKYPEKSLKINSKNSFANTSLASWYLYAPRIAGGDPNKTLSFAQLGFKYGQTDVEKY
        FANIWISQAHFLLKNEKESLKYILKAGEIFPNGAFHKIVLEQNKAGNLFMDFPIKN"
gene    17388..18170
        /gene="BBJ24"
CDS     17388..18170
        /gene="BBJ24"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ24"
        /db_xref="PID:g2690212"
        /translation="MLRALLIFLVNISLFSFEKEVKVYIDKIESVHTKYCSGNFEFDF
        FSPDKIFTNELQNIENVILMKYRRESIQYNVLNLLMSLVLCDVSYLINDPHKYNDLIQ
        KLIRNYNCALKISLEDDNVPADYFRALGELAINLIPHNRKGLYSYFVNAKRHLETALK
        IDGDNVKAFIPLSIITVRVSNRDFYKILFAKSYIDRAEDSNLNDRQKYLKELVKSSF
        LIRTNRRLEAIECLKKATATAIFPNGNMAVLAIEKLKEGNSFYY"
gene    18202..19254
        /gene="BBJ25"
CDS     18202..19254
        /gene="BBJ25"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11

        /product="B. burgdorferi predicted coding region BBJ25"
        /db_xref="PID:g2690213"
        /translation="MKVFMKKFLFLLPCFGVFANELNDELGDFVIRGVDFEFRLDYL
        SVPNNFENNFDFILNKENDSKISPFFRLGTDYSKMFLFGTGLAYDFRKFFSKVFYEI
        RVPFIFDSKNIEHIGNFEVGYNFDYLRLENRFRSGLMNHLIKETEANVGGSYHNALTV
        ENTVAILLPIYYSEFQRADIRTSFLYKYLSDNNEQFYRVHWNLKYLVSIPFGELGFKA
        DLGVAGDFKKSSSSIFETGFDYNALNFYALTIPKMGQDSLYFNVVSNFGLEYRLFFLE
        SLKNLASDLFLVLSADIGYGIKEDLLLDKGKFLYILGFGMGYKLFKEVPFVFKVGINQ
        DKKLSFGFLLSSIIFE"
gene    19313..20008
        /gene="BBJ26"
CDS     19313..20008
        /gene="BBJ26"
        /note="similar to GB:L77117 SP:Q58206 PID:1591493 percent
        identity: 35.94; identified by sequence similarity;
        putative"
        /codon_start=1
        /transl_table=11
        /product="ABC transporter, ATP-binding protein"
        /db_xref="PID:g2690182"
        /translation="MGNVNLDLKLVNKKYKIGQEITHANKDISLNLKSRDMVWISGPT
        GSGKTLMNLLSGIDSLDTGEIRFNSTLLSSMNEKDRTLFRRYNVGLIFQHFELIPSL
        TGFDNISLPLRFSRESAKRLKSKAEELIEFFKLSKFVNKKPRYMSGGQRQRIGIARAF
        VYDPKLIIGDEITSHLDKETAIFVYTSIQKYLKEKNAIGIFVSHDYNLKNLANKLYRI
        EDGVLSLVGGECV"
gene    19995..21230
        /gene="BBJ27"
CDS     19995..21230
        /gene="BBJ27"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ27"
        /db_xref="PID:g2690214"
        /translation="MSVFKLAFYNIFRDLRRTIILSLLLASSVVFLLVFVGYMNFSRE
        GMEKSFVSSSGHQIAKENYFNPKFSNLKNGLLLEEKDILRNEIDSYDELQSTNLI
        VNFDGLLGNSSTSNPVFAFAYEDPDIITSSLSLLEGEPIFHDSNAGEFLLGSNLASSF
        GIEKITETNSDLTLMTNLLGRGLNFQNIKVAGIIKPFSTADNIFAITTIKTLKDLFA
        FEGGAHVIQVFLKDSSTLETFKKKLDNFKKNKGISFDYNDWFEINPYFKSVLGMTRT
        FMFILVLISLLFIFARFQIMTALSIERTRELGTLRAIGLTKLELFYSLFLEIVIISVV
        NIVVGVILAYFAKLFIQFQKISFTPPGYSETYYINIFYYASDIIYVSIFMLILAIFSS
        ILPFSKASKKSVVEVMNDA"
gene    21220..21948
        /gene="BBJ28"
CDS     21220..21948
        /gene="BBJ28"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ28"
        /db_xref="PID:g2690194"
        /translation="MMLKIFVIVFNFCVLNLLNAGDGKSLIKEFENLYYPQLKNGIYA
        FKMNFKINVKNNLEESVGLRIINVDNKDVRLJYMSGSKTDFAFLSIRNKGHFMLGRQA
        KIPIKVSSSYKVKGASELKDIIGLSFNTDFVLKSEDNRVEFQSKEKSIYPPVDLLKI
        NKNDFKTLHKDKKLKILKEVIYRKGNIKGIDAFVFEIEDKAFNDSSTKIIVENIIST
        NLNNSIFSLKGFNRIFDLYSRYIN"
gene    21971..23008
        /gene="BBJ29"
CDS     21971..23008
        /gene="BBJ29"
        /note="hypothetical protein; identified by Glimmer;
        putative"
        /codon_start=1
        /transl_table=11
        /product="B. burgdorferi predicted coding region BBJ29"
        /db_xref="PID:g2690195"
        /translation="MVFFKFCLILFLALPGFLFAESSLEFLKEHLGFKTKFSLLLPDE
        DKKDFFGTGALQFDNEAYLGFLLNKILEFGIAPSFIVQNNDQYFSFNKLFFNLSFND
        FIFKLGRQNYILGNGLLIENIVLKRTIEPEWFEFYYFISNYSVSLGSMLDKESLDKF
        SSPKYLSPWLYFQASFNEVDLLAMVFEPFSIENKFDITVIFDLFEIYSGIFFYSTL
        RQDLLWSKNYKFDDDDNRFLLGLRYYISFENDVFNDLSIVEFESYSKNNNYFLGAGIKL
```

```
                AWIYELQQTTVSLRTNLNTHSLQFYFENNFLILKECSLKVSNILEFNKKINLKDTPFY
                NMFSIELKIEL"
gene            23018..23122
                /gene="BBJ30"
CDS             23018..23122
                /gene="BBJ30"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ30"
                /db_xref="PID:g2690196"
                /translation="MVYYIFSCYIQDECIMKLGADKIIRQKDRYCLKT"
gene            23366..24088
                /gene="BBJ31"
CDS             23366..24088
                /gene="BBJ31"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ31"
                /db_xref="PID:g2690197"
                /translation="MEKGGGHSNPKVVGVKGMFESALEIKALSKEDNPKDMVEKASIS
                AKESKDKVLSSKKGNASIALGEVPSLVRDIRESAKKIDESLKLLIASGYDASYPLSDN
                MKIGIGRICLLDKILEAVMPKINDSDNDFDDNSCEKIKKVDVFEFNKKNKDEYSCSIM
                SLNSYPKTKGKDLVKKCINELMKDVEKYLGDVKGNGNGVCGELLQGAKSASGQFGESL
                KLLSEAASSIAEACKRLAYDIL"
gene            complement(24386..24517)
                /gene="BBJ32"
CDS             complement(24386..24517)
                /gene="BBJ32"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ32"
                /db_xref="PID:g2690198"
                /translation="MLNSNSLKADINIIVRRESTILILLKGLDYSISALFVLKRFMV"
gene            24681..24794
                /gene="BBJ33"
CDS             24681..24794
                /gene="BBJ33"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ33"
                /db_xref="PID:g2690199"
                /translation="MLLCVNVILNNFSLVILKEQLFNKKSNVNVYSESSGC"
gene            complement(25337..26407)
                /gene="BBJ34"
CDS             complement(25337..26407)
                /gene="BBJ34"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ34"
                /db_xref="PID:g2690200"
                /translation="MIKGNTFILILVTTMFVSCKFYGSDDTNKKNTSLNGDTREIDNI
                GSVILEQDGNKKGDTTASKVALDQVTEHANSELMLNDDPDSSISKYNQENTTGKLTEE
                DMDKLRAFFVKFITYQGILNSIYNKTRSYNTIATYSGCAANYNSIGCFSEQPSARRSQ
                ALNDLEKNKLDEYTKLNQMLKETTQDYCPKALDNAIEEYKRAITIAKEAEDKIKKIT
                SFTIDEGNNEERKENVDNLKKVNNILSISEKTIETASVAYANAFAVIVSRLSSAEFI
                TAVNEFKDATEKYANGNKGDHAVDVVVGAIAGIAFDNENRFERAKMFANKEKGAEVDK
                MIAAIEKLRATYTAVKPKNKDK"
gene            26858..26962
                /gene="BBJ35"
CDS             26858..26962
                /gene="BBJ35"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ35"
                /db_xref="PID:g2690201"
                /translation="MKVCERREEYKINVLFKDSSIIDKENYIEYYRT"
gene            complement(26995..28053)
                /gene="BBJ36"
CDS             complement(26995..28053)
                /gene="BBJ36"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ36"
                /db_xref="PID:g2690202"
                /translation="MKKKSNICISLLVTILFVSCKFFGNKSASKEKEETSFSDTASKI
                SKSGTAASSDKQEKNTSDVTGDAKKHTSPYMLADALIVSDTTNRDRKQENKDKLNE
                EDKKKLNAFFSTKTKYQSSLDSIYNKYTGYNTIDTYGSCDTYRIECFSVGPSEKRKQ
                ALADLEKLKLDEKYTQLSTMLKSAVPSYYKKNLDDSIAQYKEAIKQAIEAESKIETVK
                DYATAQSAADDEKKRNIDNLKIVRDVLLIIKKTIEKASRSYADAFAIATSSLSCSEFK
                QAVKEFNDAAKQYANGNKGDNAVNVIVGTISSMPYVKFKDEFARAKMFARNYRGDEVD
                KMIRAIDKLCDVYKKVAL"
gene            complement(28278..28442)
                /gene="BBJ37"
CDS             complement(28278..28442)
                /gene="BBJ37"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ37"
                /db_xref="PID:g2690203"
                /translation="MILLLLPYLVVNLGKLLKILVFAAKEYANGKGKIMLMLLGA
                ISSMAYNEFE"
gene            complement(28605..28703)
                /gene="BBJ38"
CDS             complement(28605..28703)
                /gene="BBJ38"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ38"
                /db_xref="PID:g2690204"
                /translation="MALLVNFSKMLKDDVPGYKDNQQYFRECNCKV"
gene            complement(29264..29401)
                /gene="BBJ39"
CDS             complement(29264..29401)
                /gene="BBJ39"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ39"
                /db_xref="PID:g2690205"
                /translation="MKISGIVVLGMITFSVISELISPRQFYSLQEFCSQLYNIIKKLK
                K"
gene            29827..29922
                /gene="BBJ40"
CDS             29827..29922
                /gene="BBJ40"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ40"
                /db_xref="PID:g2690215"
                /translation="MYEHKKEKLILFLKYKVQHIIFLFRLISLV"
gene            complement(29905..30771)
                /gene="BBJ41"
CDS             complement(29905..30771)
                /gene="BBJ41"
                /note="similar to GP:1553115 percent identity: 25.93;
```

```
                    identified by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="antigen, P35, putative"
                    /db_xref="PID:g2690183"
                    /translation="MKNLKLNIIKLNVITALITSICISCAPFGNVNPEKNPTTSKS
                    LKKTKRSNNSRLKNTSNHTNSENLTGNSTKNPSENNQNLENESQNSKSSNQNSQET
                    TISKLKNIGKDLEAQKKKEDTGITKMSKIDNAKYDFLETFKLKQDDVFMFXAKMLKR
                    IIYPSLNYDTKKILVLKEILEKLDTEDNNRRIAGQFLETSRDIQLHLEDTYLKKIQDT
                    LQTLSEKEAEKLQGVKLDLKKKQNFAKSLNATIDAYNKNVDNIKIDNKALAKHIKDK
                    YSHPLYLLNQAD"
gene                complement(30942..31133)
                    /gene="BBJ42"
CDS                 complement(30942..31133)
                    /gene="BBJ42"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ42"
                    /db_xref="PID:g2690216"
                    /translation="MRCAKTLTANNIESYNQNFGDIKTNIKHFEDHFMKEYFDFILRI
                    KSIKSIQIVVLDFLKREEL"
gene                31220..32140
                    /gene="BBJ43"
CDS                 31220..32140
                    /gene="BBJ43"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ43"
                    /db_xref="PID:g2690217"
                    /translation="MLPDEDKKDFFGTGALQFANEAYLGFLLNYKILEFGIAPSFIVQ
                    NNDQYFSFNKLFFNLSFNDFIFKLVRQNYLGNGLIENIVLKRTIEPEWFEFYFI
                    SNYSVSLGSMLDKESLDKFLSPKYLSPWLIFQASFNEVDLLAMVEPPFSIENKTFDIA
                    VIFDFLFEIYSGIFFYSTLRQDLLWSKNYKFDDDDNRFLLGLRYYISFENDVFNDLSI
                    VFEESYSKNNNYFLGAGIKLAWIYELQTTVSLRTNLTHSLQFYFENNFLLIKECSLK
                    VSNILEFNKKINLKDTPFYNMFSIELKIEL"
gene                32150..32254
                    /gene="BBJ44"
CDS                 32150..32254
                    /gene="BBJ44"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ44"
                    /db_xref="PID:g2690218"
                    /translation="MAYYIFSCYIQDECIMKLGADKIIGQKDRYCLKT"
gene                32498..33220
                    /gene="BBJ45"
CDS                 32498..33220
                    /gene="BBJ45"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ45"
                    /db_xref="PID:g2690219"
                    /translation="MEKGGGHSNPKVVGVKGMFESALEIKALSKEDNPKDMVEKASIS
                    AKESKDKVLSSKKGNASIALGEVPSLVRDIRESAKKIDESLKLLIASGYDASYPLSDN
                    MKIGIGRICLLDKILEAVMPKINDSDNDNFDDNICERIKKVVDEFNKNEKDEYSCCSIM
                    SLNSYPKTKGKDLVKKCINELMKDVEKYLGDVKGNGNVCGELLQGAKSASGQFGESL
                    KLLSEAASAIAEACKRLAYGIL"
gene                complement(34372..34668)
                    /gene="BBJ46"
CDS                 complement(34372..34668)
                    /gene="BBJ46"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
```

```
                    /product="B. burgdorferi predicted coding region BBJ46"
                    /db_xref="PID:g2690220"
                    /translation="MDKFLTSNHPPIIFTIGALCATVLICLIIFIIHGIINPILIK
                    KFKSINNSLQKTTKEFEMKKRVGQLESISKKLELNNKPAKKPISLYGLKKSQI"
gene                complement(34905..35591)
                    /gene="BBJ47"
CDS                 complement(34905..35591)
                    /gene="BBJ47"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ47"
                    /db_xref="PID:g2690221"
                    /translation="MRNISNCIKYIILTMLIGLLIFCCATFWLIGIFYSNNFKEERN
                    YSISPIDSVIMRKCYFKEFKSGLIKSVFFKLDVNVNSKNFKELNKVDKQNLNSYPS
                    YHMEFVVVDNGFLMNFKNVIFNGIDDAKLYDQRDMVVGGFRYSKEAYFQIIGNYDVKL
                    NMKKQYTPAIVVNVFKININDALFNSLLKQKTLKVTLISHNNKEYILQTNNFLSKYNF
                    QTPEKENSSY"
gene                complement(35634..36272)
                    /gene="BBJ48"
CDS                 complement(35634..36272)
                    /gene="BBJ48"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ48"
                    /db_xref="PID:g2690222"
                    /translation="MLYLFTTKMKIKFLTISLAILLANLIILLNLVLFIVNTSTSSP
                    YIVPSEKIDILHQSNTGAVKFKISLINHLGSVAIVYDYNSASSRFYLDFEIVTNKKPF
                    NLLDVSLNDVVIKPEVLLASNSKLRFEEGQYYLNFDDSIEKTGFFVDLDRNEYLNLA
                    EIARISGINFRVKCIERETGVLRNVLFKLSVEKGKKFFDLIERYNNNIGKVS"
gene                complement(36312..36455)
                    /gene="BBJ49"
CDS                 complement(36312..36455)
                    /gene="BBJ49"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ49"
                    /db_xref="PID:g2690223"
                    /translation="MLLIHLLLMNLFELAFIFASNKKGSEVDTMITAIDKLITIYKK
                    VKS"
                    36502..37206
                    /note="This region contains an authentic point mutation,
                    causing a premature stop, and is not the result of a
                    sequencing artifact.; similar to GB:L31615 GB:L31418
                    PID:520783 PID:551742 percent identity: 35.58; identified
                    by sequence similarity; putative"
                    /gene="BBJ50"
gene                complement(37415..37917)
                    /note="This region contains an authentic frame shift and
                    is not the result of a sequencing artifact.; similar to
                    GP:2039290 percent identity: 65.62; identified by sequence
                    similarity; putative:vlsE1 protein (vlsE1), authentic
                    frameshift"
                    /gene="BBJ51"
gene                complement(37927..38522)
                    /note="This region contains an authentic frame shift and
                    is not the result of a sequencing artifact.; similar to
                    PID:558831 percent identity: 29.10; identified by sequence
                    similarity; putative:vlsE1 protein, authentic frameshift"
                    /gene="BBJ52"
BASE COUNT      13739 a    4828 c    5294 g   14966 t     2 others
ORIGIN

Query Match              4.7%;   Score 47.2;   DB 2;   Length 38829;
Best Local Similarity   50.4%;   Pred. No. 1.6;
Matches  115;   Mismatches  113;   Conservative  0;   Indels  0;   Gaps  0;
```

```
Qy    769 aaggagattgatctgacttcgactggatccacttaactatatctaattttacg 828
          |||||||| || |||  ||  |||  || ||||||||| |||| |  ||||
Db  17722 AAAGCACAATTATAATTGCGAATTAGTTTTGAATCAAATCATTGTATTATGTGATCA 17663

Qy    829 gcattaaggtttgtaaattggtcgacagcttagtcaaaactacttatgcactttaatat 888
          | ||||||| || || |  || || |  ||||||||| ||| || |  ||||
Db  17662 TTAATTAAGTAAGAAACGTCACACAACAAAGACATAGGTAAAATTAAGTAATAATTATAT 17603

Qy    889 ggcttcttgtctatatttatttatttacatggctgtatctaggttgcatttaagat 948
          |||||||||| ||| || |||| |||||| |||||| ||| |||||||
Db  17602 TGAATACTTTCCCTTCCTATATTTCATTAAAATAACACATTTTCTATGTTTGAAGCTCATT 17543

Qy    949 ttagtaccttgtcagattaaaagaaacgaaagtaaattaaaaaaa 996
          ||  ||  |||| || |||||||  |||  ||||||  ||| |||
Db  17542 GTAAAAATTTTATCGGGTGAAAAAAATCAAACTCAAAATTACCAGAA 17495
```

```
RESULT    8
AE000787/c
LOCUS        AE000787      38829 bp      DNA               BCT        10-DEC-1997
DEFINITION   Borrelia burgdorferi plasmid lp38, complete plasmid sequence.
ACCESSION    AE000787
NID          g2690175
KEYWORDS     .
SOURCE       Lyme disease spirochete.
ORGANISM     Borrelia burgdorferi
             Eubacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
             burgdorferi group.
REFERENCE    1 (bases 1 to 38829)
AUTHORS      Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
             Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
             Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
             Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
             Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
             Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
             Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
             Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
TITLE        Genomic sequence of a Lyme disease spirochaete, Borrelia
             burgdorferi
JOURNAL      Nature 390 (6660), 580-586 (1997)
MEDLINE      98065943
REFERENCE    2 (bases 1 to 38829)
AUTHORS      Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
             Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
             Gwinn,M., Dougherty,B., Tomb,J.-F., Fleischmann,R.D.,
             Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
             Salzberg,S., Hanson,M., van-Vugt,R., Palmer,N., Adams,M.D.,
             Gocayne,J.D., Weidman,J., Utterback,T., Watthey,L., McDonald,L.,
             Artiach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
             Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
TITLE        Direct Submission
JOURNAL      Submitted (08-DEC-1997) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..38829
                     /organism="Borrelia burgdorferi"
                     /db_xref="taxon:139"
                     /plasmid="lp38"
     gene            482..667
                     /gene="BBJ01"
     CDS             482..667
                     /gene="BBJ01"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ01"
                     /db_xref="PID:g2690206"
                     /translation="MKYHIIVSIFVFLFLNACNPDSNTNQNNSKKELKTGRIPNKQIK
                     NALDDLKNLIETASAG"
     gene            927..1211
                     /gene="BBJ02"
     CDS             927..1211
```

```
                     /gene="BBJ02"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ02"
                     /db_xref="PID:g2690207"
                     /translation="MINHLLCLKKETLDTLTVSDLKKLKNSFEKLLSIKTDISKMLEQ
                     LLDYKDYINSTKTDIAKLESHLTELYKQIVKKSFQTTELKNNILSISNL"
     gene            1593..1745
                     /gene="BBJ03"
     CDS             1593..1745
                     /gene="BBJ03"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ03"
                     /db_xref="PID:g2690184"
                     /translation="MAKKHPYYYNFLNVSLTLGIGSFWPRGLYWQHGELLRSQVL
                     GEIYLR"
     gene            complement(2268..2381)
                     /gene="BBJ04"
     CDS             complement(2268..2381)
                     /gene="BBJ04"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ04"
                     /db_xref="PID:g2690185"
                     /translation="MLLYKTFSMIGFCFRALSENEKHFLMLLFSAGTIF"
     gene            complement(2765..3261)
                     /note="This region contains an authentic frame shift and
                     is not the result of a sequencing artifact.; similar to
                     GP:2228714 percent identity: 94.44; identified by sequence
                     similarity; putative:transposase-like protein, authentic
                     frameshift"
     gene            3486..3632
                     /gene="BBJ05"
     gene            3486..3632
                     /gene="BBJ06"
     CDS             3486..3632
                     /gene="BBJ06"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ06"
                     /db_xref="PID:g2690186"
                     /translation="MISSKSFIIFLLSYLENYCVLFPFLIYLKNYMQSLSQAGHYK
                     ELNY"
     gene            complement(4164..4307)
                     /gene="BBJ07"
     CDS             complement(4164..4307)
                     /gene="BBJ07"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ07"
                     /db_xref="PID:g2690187"
                     /translation="MEMNLLINRFIFILNIFFDDINRGVPLLYRDKAIASRPKIRFSA
                     ILI"
     gene            4576..5496
                     /gene="BBJ08"
     CDS             4576..5496
                     /gene="BBJ08"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="B. burgdorferi predicted coding region BBJ08"
                     /db_xref="PID:g2690188"
```

```
gene
            /translation="MFLYTLLTIGLISCNLDSKLPNKEQKNNNDIKETLGSSVQENAL
            NNLYGNQEEKKDFKNFEELKDESLIALAKSLASTRPTTVGNIESAVLPVGHVVSLETS
            ANKVSIPTISIKHNQKKEIKKEDLSPSTKEEKKADKAIKDIENLIRDSGFPELIESMY
            SLKHEYTLIRNNFYDVITKIRNNKTSLIKNGRNNRDKIKELTQLGNNLKIVDELDEIM
            VHIDIAEQEIRSAAFFNEAKEILKEGIIKRLESENKVASQLARQALNKVEDALKSLE
            ASSKRGLAMGRRIIKELIENAKTVLSKS"
CDS
            6089..6862
            /gene="BBJ09"
            6089..6862
            /gene="BBJ09"
            /note="similar to    percent identity: 99.61; identified by
            sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="outer surface protein D (ospD)"
            /db_xref="PID:g2690176"
            /translation="MKKLIKILLSLFLLSISCVHDKQELSKSNLNNQKGYLDNEG
            ANSNYESKKQSILSELNQLLKQTNSLKEAKNTDNLNASNEANKVEAVINAVNLIS
            SAADQVKSATKNHDLAQMAEIDLEKIKNSSDKAIFASNLAKEAYSLTKAAEQNMQKL
            YKEQQKISESESESDYSDSAEIKQAKEAVEIAWKATVEAKDKLIDVENTVKETLDKIK
            TETTNNTKLADIKEAAEEVLQIAKNAKEIVQEVVALNT"
gene
            7270..7476
            /gene="BBJ10"
            7270..7476
            /gene="BBJ10"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BBJ10"
            /db_xref="PID:g2690189"
            /translation="MGARDEFKNGIRKLDEKIERARIELSTKITDVRSELSKQVKGLI
            FPFYWILGIFIPAVIGLFLYLLQK"
CDS
            complement(7780..7965)
            /gene="BBJ11"
            complement(7780..7965)
            /gene="BBJ11"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BBJ11"
            /db_xref="PID:g2690190"
            /translation="MFLSFCNLCITKNIHNIGSLRRVYTNFVVDYAFWLSKIAEDYFP
            DAKDIVCIRGIKKPYI"
gene
            complement(8633..8725)
            /gene="BBJ12"
            complement(8633..8725)
            /gene="BBJ12"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BBJ12"
            /db_xref="PID:g2690191"
            /translation="MVLGIWKKAENYVDKFDELIIVKNKELFVI"
CDS
            complement(8877..9155)
            /gene="BBJ13"
            complement(8877..9155)
            /gene="BBJ13"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BBJ13"
            /db_xref="PID:g2690192"
            /translation="MLHINHANEAFKEISSAKDKIDSSDIGKGIVDVQKTHNRLHDLL
            RDIKTEFYSQKNSPLNGVRAEKSKKKNQILFRSTSSINGKIRNANYIL"
gene
            9125..9286
            /gene="BBJ14"
            9125..9286
            /gene="BBJ14"
            /note="similar to PID:62180 percent identity: 33.33;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2690177"
            /translation="MLHLHDLYATNFVIFQLSIYNLFVTFFKFSSNAEPDPLLVPFDL
            LLFKDSSKP"
gene
            complement(10040..10168)
            /gene="BBJ15"
            complement(10040..10168)
            /gene="BBJ15"
            /note="hypothetical protein; identified by Glimmer:
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BBJ15"
            /db_xref="PID:g2690193"
            /translation="MSISWLFIIFKFISQTIDNVQKININKTGVFVALIKLLKLNI"
CDS
            complement(10518..11105)
            /gene="BBJ16"
            complement(10518..11105)
            /gene="BBJ16"
            /note="similar to GB:U03641 PID:458218 percent identity:
            35.19; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2690178"
            /translation="MKIAEKDKEALFSNRFGDCNEETNLNDDQDKELANYNNLKEQL
            KYNLKDDINNKIQRMKILYEIKQKELYKYDGFARFNDFIKSFEVAKSQAYRYLKIYQK
            VLEGKVSIDKIKEVGFKAILRDIKAKDSLNEDNHSESEGANESIPIRILVDKELYNF
            CKQDTKRLYFIIEKIYKEKRDVLSELIIEYEKNKK"
gene
            complement(11152..11889)
            /gene="BBJ17"
            complement(11152..11889)
            /gene="BBJ17"
            /note="similar to GP:2182756 percent identity: 39.43;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2690179"
            /translation="MDRKKSNIITIANLKGGVGKSTLSLFSYVLKDLGKKVLLIDMD
            SQNALTSYFRKYVNFDKNNIYNLLIGNVFDQCISKINDNIFIIPSHPFLDEFNDKN
            LDNKENLLSFCLDKNVLGHDFDYIFLDTPPSFSFILKNALNTTNHIIIPVQPETWSIE
            SLEILMKKITDKSYNISIVNQFIKNRNILKEVEDALYKYRSNYIKGKIHYNSIKVF
            IINRLEPDIKSKYYKEAKDVLKNIIDL"
CDS
            complement(11862..12452)
            /gene="BBJ18"
            complement(11862..12452)
            /gene="BBJ18"
            /note="similar to GB:U03641 PID:458217 percent identity:
            39.38; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2690180"
            /translation="MNDLLEKLKARKKEIISKKEAEYNNNINKGTKERTAFFRIEEID
            NKKIYTKIFKLYLVKFRIANKDNKLSLTFQKLNKKNYYLFNLFPLKEDNKFLGIKYG
            WDKLEKPFFLRQNNKSYVIKKLYLEFKFSKGSIKCYVQSLRTLLRKKDKESTKYKF
            NLEHIKRMENTVYKFYSKKLKNKGVYKWIEKNQIL"
gene
            complement(12445..13440)
            /gene="BBJ19"
            complement(12445..13440)
            /gene="BBJ19"
            /note="similar to GP:1655797 percent identity: 27.04;
            identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /db_xref="PID:g2690181"
            /translation="MALNLLNHHNHKTNTNKLLSKDSRLKKIISVINYLNKSFEKKYD
            ISIYRIHFNSEKLKELYPHQIDILRVLNSNISKEGYKPTVIRTLREDLRFLIHIKAI
```

```
                EKKILTFSNNLGFKGKLCIYKVSPIAVKLISAYFEAYKADLYRKIKKSKDGLDTQNV
                TKNVTVINYHKNIYNKNSIETVFKKSYTKKKTKKKNKKET KNNLEKRLKLPEEITK
                EISIAKKTKNPDKTYKNTLFNYKDFLNYLSYDYKKEDISYFFLSKLKEYKNKIHFMR
                KYAPYKTDFYLLAGEFKDSYHSKWKTNFSGHVKEIANNLSKILEKELKFE"
     gene       complement(13772. .13936)
                /gene="BBJ20"
     CDS        complement(13772. .13936)
                /gene="BBJ20"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ20"
                /db_xref="PID:g2690208"
                /translation="MQRAKHTDRTNLENLLNSFNKMKNSNIHVQNKTTRNRKEIGSP
                NYITKTIKVY"
     gene       15514. .15660
                /gene="BBJ21"
     CDS        15514. .15660
                /gene="BBJ21"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ21"
                /db_xref="PID:g2690209"
                /translation="MARAKIYLHISFKYVLKKPILVLIKPRKLSPMFLFILKKKQKAC
                LYIF"
     gene       complement(15902. .16003)
                /gene="BBJ22"
     CDS        complement(15902. .16003)
                /gene="BBJ22"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ22"
                /db_xref="PID:g2690210"
                /translation="MELLKMKKVEYVLYSKDKHTILLDANILQGKS"
     gene       16505. .17329
                /gene="BBJ23"
     CDS        16505. .17329
                /gene="BBJ23"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ23"
                /db_xref="PID:g2690211"
                /translation="MNGVIMREISCCFLLLTFSVVCVYSFDVSSRKFYGILEGYYSGK
                IEELSKKNDEDVYIYRFGKFKETLSEMSSGIKSYFFNLVDYQIARLLQNKEGRRNSSK
                SYSVLKSTQKSLELIASTDFKGLEKTIQSDICRILGDVNLMLLRYAGGAALSKLANE
                ARKYFEKSLKINSKNSFANTSLASWYLYAPRIAGGDPNKTLSFAQLGFKYGQTDVEKY
                FANIWISQAHFLKNEKESLKYILKAGEIFPNGAFHKIVLEQNKAGNLFMDFPIKN"
     gene       17388. .18170
                /gene="BBJ24"
     CDS        17388. .18170
                /gene="BBJ24"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ24"
                /db_xref="PID:g2690212"
                /translation="MLRALLIFLVNISLFSFEKEVKVYIDKIESVHTKYCSGNFEPDF
                FSPDKIFTNELQNIENVILMKYRRESIQYNLNLMSLVLCDVSYLINDPHKYNDLIQ
                KLIRNYNCALKISLEDDNVPADYFRALGELAINLIPHNRKGLYSYFVNAKRHLETALK
                IDGDNVKAFIPLSILYLYVRVSNRDFYKILFAKSYIDRAEDSNLNDRQKYLKELVKSSF
                LIRTNRRLEAIECLKKATAIFPNGNMAVLAIEKLKEGNSFYY"
     gene       18202. .19254
                /gene="BBJ25"
     CDS        18202. .19254
                /gene="BBJ25"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ25"
                /db_xref="PID:g2690213"
                /translation="MKVFMKKFLFLILPCFGVFANELNDELGDFVIRGVDFEFRLDYL
                SVPNNFENNFDFILNIKENDSKISPFFRLGTDYSKMFLFGTGLAYDFRKFFSKVFYEI
                RVPFIFDSKNIEHIGNEFEVGYNFDYLRLENRFSRGLMNHLIKETEANVGGSYHNALTV
                ENTVAILLPIYYSEFQORADIRTSFLYKYLSDNNEQFYRVHWNLKYLVSIPFGELGFKA
                DLGVAGDFKKSSSSIFETGFDYNALNFYALTIPKMGQDSLYFNVVSNFGLEYRLFFLE
                SLKNLASDLFLVLSADIGYGIKEDLLLDKGKFLYILGFGMGYKLFKEVPFVFKVGINQ
                DKKLSFGFLLSSIIFE"
     gene       19313. .20008
                /gene="BBJ26"
     CDS        19313. .20008
                /gene="BBJ26"
                /note="similar to GB:L77117 SP:O58206 PID:1591493 percent
                identity: 35.94; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein"
                /db_xref="PID:g2690182"
                /translation="MGNVNLDKLIVNKKYKIGQEIIHANKDISLNLKSRDMWISGPT
                GSGKTTLMNLLSGIDSLDTGEIRFNSTLLSSMNEKDRTLFRRYNVGLIFQHFELIPSL
                TGFDNISLPLRFSRESAKRLKSKAEELIEFFKLSKFVNKKPRYMSGGQRQRIGIARAF
                VYDPKLLIGDEITSHLDKETAIFVYTSIQKYLKEKNAIGIFVSHDYNLKNLANKLYRI
                EDGVLSLVGGGECV"
     gene       19995. .21230
                /gene="BBJ27"
     CDS        19995. .21230
                /gene="BBJ27"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ27"
                /db_xref="PID:g2690214"
                /translation="MSVFKLAFYNIFRDLRRTIILSLLLASSVFLLVFVGYMNFSRE
                GMEKSFVSSSGHQIAKENYFNPKFSNLKNGLLLEEKDINLIRNEIDSYDELQSTNLI
                VNFDGLLGNSSTSNPVFAFAVEDPDIITSSLSLLEGEPIFHDSNAGEFLLGSNLASSF
                GIEKITETNSDLTLMNLLGRGLNFQNIKVAGIIKPFPSTADNIFAITTIKTLKDLFA
                FEGGAHVIQVFLKDSSTLETFKKLDNFKNKGISFDYNDWFEINPYFKSVLGMTRTT
                FMFILVLISLLIFIAFFQIMTALSIERTRELGTLRAIGLTHLKLEFYSLFLEIVIISVV
                NIVVGVILAYFAKLFIQFQKISFTPPGYSETYYINIFYYASDIIYVSIFMLLIAIFSS
                ILPFSKASKKSVVEVMNDA"
     gene       21220. .21948
                /gene="BBJ28"
     CDS        21220. .21948
                /gene="BBJ28"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ28"
                /db_xref="PID:g2690194"
                /translation="MMLKIFVIVFNFCVLNLLNAGDGKSLIKEFENLYYPQLKNGIYA
                FKMNFKINVKNNLEESVGLRIINVDNKDVRLIYMSGSKTDFAFLSIRNKGHFMLGRQA
                KIPIKVSSSYKVKGASELKDILGLSNTDFVLLKSEDNRVEFQSKEKSIYPFVDLKI
                NKNDFKTLHKDKKLKIKEVIYRKGNIKGIDAFVYFEIEDKRAFNDSSTKIYVENIST
                NLNNSIFSLKGFNRIFDLYSRYIN"
     gene       21971. .23008
                /gene="BBJ29"
     CDS        21971. .23008
                /gene="BBJ29"
                /note="hypothetical protein; identified by Glimmer;
                putative"
                /codon_start=1
                /transl_table=11
                /product="B. burgdorferi predicted coding region BBJ29"
                /db_xref="PID:g2690195"
                /translation="MVFFKKFCLILFLALPGFLFAESSLFKEHLGFKTKFSLLLPDE
```

```
                    DKDFFGTGALQFDNEAYLGFLLNYKILEFGIAPSFIVQNNDQYFSFNKLFFNLSFND
                    FIFKLGRQNYYLGNGLIENIVLKRTTIEPEWFFEFYYFISNYSVSLGSMLDKESLDKF
                    SSPKYLSPWLYFQASFNEVDLLAMVEFFPSIENKTFDITVIFDFLFEIYSGIFFYSTL
                    RQDLLWSKKNYKFDDDNRFLGLRYYISFENDVNDLSIVFESSKNNYFLGAGIKL
                    AWIYELLQTTVSLRTNLNTHSLQFYFENNFLIKECSLKVSNILEFNKKINLKDTPFY
                    NMFSIELKIEL"
gene                23018..23122
                    /gene="BBJ30"
CDS                 23018..23122
                    /gene="BBJ30"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ30"
                    /db_xref="PID:g2690196"
                    /translation="MVYYIFSCYIQDECIMKLGADKIIRQKDRYCLKT"
gene                23366..24088
                    /gene="BBJ31"
CDS                 23366..24088
                    /gene="BBJ31"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ31"
                    /db_xref="PID:g2690197"
                    /translation="MEKGGGHSNPKVVGVKGMFESALEIKALSKEDNPKDMVEKASIS
                    AKESKDKVLSSKKGNASIALGEVPSLVRDIRESAKKIDESLKLLIASGYDASYPLSDN
                    MKIGIGRICLLDKILEAVMPKINDSDNDNFDDNSCEKIKKVVDEFNKNEKDEYSCSIM
                    SLNSYPKTKGKDLVKKCINELMKDVEKYLGDVKGNGNGVCGELLQGAKSASGQFGESL
                    KLJSEAASSIAEACKRLAYDIL"
gene                complement(24386..24517)
                    /gene="BBJ32"
CDS                 complement(24386..24517)
                    /gene="BBJ32"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ32"
                    /db_xref="PID:g2690198"
                    /translation="MLNSNSLKADINIIVRRESTILILKGLDYSISALFVLKRFMV"
gene                24681..24794
                    /gene="BBJ33"
CDS                 24681..24794
                    /gene="BBJ33"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ33"
                    /db_xref="PID:g2690199"
                    /translation="MLLCVNVILNNFSLVILKEQLFNKKSNVNYSESSGC"
gene                complement(25337..26407)
                    /gene="BBJ34"
CDS                 complement(25337..26407)
                    /gene="BBJ34"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ34"
                    /db_xref="PID:g2690200"
                    /translation="MIKGNTFILILVTTMFVSCKFYGSDDTNKKNTSLNGDTREIDNI
                    GSVILEQDGNKKGDTTASKVALDQVTEHANSELMNDDPDSSISKYNQENTTGKLTEE
                    DMDKLKAFFVKTITYQGILNSIYNKYTRSYNTIATYSGCANYNSIGCFSEGPSARRSQ
                    ALNDLEKNKLDEEYTKLNQMKETTQDYCPKALDNAIEEYKRAITIAKEAEDKIKKIT
                    SFTIDEGNNNEERKENVDNLKKVNNILSISEKTIETASVAYANAFAVIVSRLSSAEFI
                    TAVNEFKDATEKYANGNKGDHAVDVVVGAIAGIAFDNENRFERAKMFANKEKGAEVDK
                    MIAAIEKLRATYTAVKPKNKDK"
gene                26858..26962
                    /gene="BBJ35"
```

```
                    26858..26962
                    /gene="BBJ35"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ35"
                    /db_xref="PID:g2690201"
                    /translation="MKVCERREEFYKINVLFKDSSIIDKENYIEYYRT"
CDS                 complement(26995..28053)
                    /gene="BBJ36"
gene                complement(26995..28053)
                    /gene="BBJ36"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ36"
                    /db_xref="PID:g2690202"
                    /translation="MKRKSNICISLLVTILFVSCKFFGNKSASKEKEETSFSDTASKI
                    SKSGTAASSDKQEKNTSDVTGDAKKHTSSPYMLADALIVSDTTNRDRDKQENKDKLNE
                    EDKKLNAFFSTTKTYQSSLDSIYNKYTGYNTIDYGSCDTYRIECFSVGPSEKRKQ
                    ALADLEKLKLDEKYTQLSTMLKSAVPSYYKKNLDDSIAQYKEAIKQAIEAESKIETVK
                    DYATAQSAADDEKKRNIDNLKIVRDVLLIIKKTIEKASRSYADAFAIATSSLSCSEFK
                    QAVKEFNDAAKQYANGNKGDNAVNVIVGTISSMPYVKFKDEFARAKMFARNYRGDEVD
                    KMIRAIDKLCDVYKKVAL"
gene                complement(28278..28442)
                    /gene="BBJ37"
CDS                 complement(28278..28442)
                    /gene="BBJ37"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ37"
                    /db_xref="PID:g2690203"
                    /translation="MILLLLPVYLVVNLGKLLKILVFAAKEYANGKGKIMLMLLGA
                    ISSMAYNEFE"
gene                complement(28605..28703)
                    /gene="BBJ38"
CDS                 complement(28605..28703)
                    /gene="BBJ38"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ38"
                    /db_xref="PID:g2690204"
                    /translation="MALLVNFSKMLKDDVPGYKDNQQYFRECNCKV"
gene                complement(29264..29401)
                    /gene="BBJ39"
CDS                 complement(29264..29401)
                    /gene="BBJ39"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ39"
                    /db_xref="PID:g2690205"
                    /translation="MKISGIVVLGMITFSVISEIISPRQFYSLQEFCSQLYNIIKKLK
                    K"
gene                29827..29922
                    /gene="BBJ40"
CDS                 29827..29922
                    /gene="BBJ40"
                    /note="hypothetical protein; identified by Glimmer;
                    putative"
                    /codon_start=1
                    /transl_table=11
                    /product="B. burgdorferi predicted coding region BBJ40"
                    /db_xref="PID:g2690215"
                    /translation="MYEHKKEKLILLFLKYKVQHIIFLFRLISLV"
gene                complement(29905..30771)
```

```
     CDS           /gene="BBJ41"
                   complement(29905..30771)
                   /gene="BBJ41"
                   /note="similar to GP:1553115 percent identity: 25.93;
                   identified by sequence similarity; putative"
                   /codon_start=1
                   /transl_table=11
                   /product="antigen, P35, putative"
                   /db_xref="PID:g2690183"
                   /translation="MKNLKLNIIKLNVITAILISICISCAPFGNVNPEPKNPTTSKS
                   LKKTKRSNNSRNLKNTSNHTNSENLTGNSTKNPSENNQNLENESQNSKSSNQNSQEET
                   TISKLKNIGKDLEAQKKKEDTGITKMSKIDNAKYDFLETFKLKQDDVFMFXAKMKLKR
                   IIYPSLNYDTKKILVLKEILEKLDTEDNNRRIAGQFLETSRDIQLHLEDTYLKKIQDT
                   LQTISEKEAEKLLQGVKLDLKKKQNFAKSLNATIDAYNKNVDNIKIDNKALAKHIKDK
                   YSHPLYLLNQAD"
     gene          complement(30942..31133)
                   /gene="BBJ42"
     CDS           complement(30942..31133)
                   /gene="BBJ42"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ42"
                   /db_xref="PID:g2690216"
                   /translation="MRCAKTLTANNIESYNQNFGDIKTNIKHFEDHFMKEYFDFILRI
                   KSIKSIQIVVLDFLKREEL"
     gene          31220..32140
                   /gene="BBJ43"
     CDS           31220..32140
                   /gene="BBJ43"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ43"
                   /db_xref="PID:g2690217"
                   /translation="MLPDEDKKDFFGTGALQFANEAYLGFLLNYKILEFGIAPSFIVQ
                   NNDQYFSFNKLFFNLSFNDFIFKLVRQNYYLGNGLIENIVLKRTTIEPEWFFEFYFI
                   SNYSVSLGSMLDKKESLDKFLSPKYLSPWLLYGASFNEVDLLAMVEFPFSIENKTFDIA
                   VIFDFLFEIYSGIFFYSTLRQDLLWSKNYKFDDDDNRFLLGLRYYISFENDVFNDLSI
                   VFESYSKNNYFLGAGIKLAWIYELLQTTVSLRTNLNTHSLQFYFENNFLILKECSLK
                   VSNILEFNKKINLKDTPFYNMFSIELKIEL"
     gene          32150..32254
                   /gene="BBJ44"
     CDS           32150..32254
                   /gene="BBJ44"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ44"
                   /db_xref="PID:g2690218"
                   /translation="MAYYIFSCYIQDECIMKLGADKIIGQKDRYCLKT"
     gene          32498..33220
                   /gene="BBJ45"
     CDS           32498..33220
                   /gene="BBJ45"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ45"
                   /db_xref="PID:g2690219"
                   /translation="MEKGGHSNPKVVGVKGMFESALEIKALSKEDNPKDMVEKASIS
                   AKESKDKVLSSKKGNASIALGEVPSLVRDIRPSAKKIDESLKLLIASGYDASYPLSDN
                   MKIGIGRICLLDKILEAVMPKINDSDNDNFDDNICEKIKKVDEFNKNEKDEYSCSIM
                   SLNSYPRKTLGDKVKKCINELMKDVEKYLGDVKGNGNVCGELLQGAKSASQFGFESL
                   KLLSEAASAIAFACKRLAYGIL"
     gene          complement(34372..34668)
                   /gene="BBJ46"
     CDS           complement(34372..34668)
                   /gene="BBJ46"

                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ46"
                   /db_xref="PID:g2690220"
                   /translation="MDKFLTSNHPPIIFTIGALCATVLICLIIFIHGIINPILIK
                   KFKSINNSLQKITKEFEEMKKRVGQLESISKKLELNNKPAKKPISLYGLKKSQI"
     gene          complement(34905..35591)
                   /gene="BBJ47"
     CDS           complement(34905..35591)
                   /gene="BBJ47"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ47"
                   /db_xref="PID:g2690221"
                   /translation="MRNISNCIKYIILIMLIGLLIFCCATFVWLIGIFYSNNFKEERN
                   YSISPIDSVIMRKCYFKEFKSGLIKSVFKLDDVVNSKNFKEKNVKDKQNLNSYPS
                   YHMEFVVYDNGFLMNFKNVIFNGIDDAKLYDRDMVVGGFRYSKEAYFQIIGNYDVKL
                   NKMKQYTPAIVVNVFKININDALFNSLLKQKTLKVTLISHNNKEYILQTNNFLSKYNF
                   QTPEKENSSY"
     gene          complement(35634..36272)
                   /gene="BBJ48"
     CDS           complement(35634..36272)
                   /gene="BBJ48"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ48"
                   /db_xref="PID:g2690222"
                   /translation="MLIFTTKMKIKFLTISLALLANLIILLNLVLFIVNTSTSSP
                   YIVPSEKIDILHQSNTGAVKFKISLINHLGSVAIVDYNSASERFYLDFEIVTNKKPF
                   NLLDVSLNDVVIKPEVLLASNSKLRFEEGQYVLNFDDSIEKTGFFVDLDLRNEYLNLA
                   EIARISGINFRVKCIERETGVLRNVLFKLSVEKGKKFFDLIERYNNNIGKVS"
     gene          complement(36312..36455)
                   /gene="BBJ49"
     CDS           complement(36312..36455)
                   /gene="BBJ49"
                   /note="hypothetical protein; identified by Glimmer;
                   putative"
                   /codon_start=1
                   /transl_table=11
                   /product="B. burgdorferi predicted coding region BBJ49"
                   /db_xref="PID:g2690223"
                   /translation="MLJILHLLMLNLFELAFIFASNKKGSEVDTMITAIDKLITIYKK
                   VKS"
     gene          36502..37206
                   /note="This region contains an authentic point mutation,
                   causing a premature stop, and is not the result of a
                   sequencing artifact.; similar to GB:L31615 GB:L31418
                   PID:520783 PID:551742 percent identity: 35.58; identified
                   by sequence similarity; putative"
                   /gene="BBJ50"
     gene          complement(37415..37917)
                   /note="This region contains an authentic frame shift and
                   is not the result of a sequencing artifact.; similar to
                   GP:2039290 percent identity: 65.62; identified by sequence
                   similarity; putative:vlsE1 protein (vlsE1), authentic
                   frameshift"
                   /gene="BBJ51"
     gene          complement(37927..38522)
                   /note="This region contains an authentic frame shift and
                   is not the result of a sequencing artifact.; similar to
                   PID:558331 percent identity: 29.10; identified by sequence
                   similarity; putative:vlsE1 protein, authentic frameshift"
                   /gene="BBJ52"
BASE COUNT     13739 a   4828 c   5294 g   14966 t      2 others
ORIGIN
```

Query Match   4.78;   Score 47.2; DB 37; Length 38829;
Best Local Similarity   50.48;   Pred. No. 1.6;
Matches 115; Conservative   0; Mismatches 113; Indels   0; Gaps   0:

Qy   769 aaggaagttagatcctgaacctcgactgggatccacttaactatatctaattttacg 828
            ||| |||||||| ||| ||||||| || | ||| |||||||||||||| | ||||
Db   17722 AAAGCACAATTATAATGTCGAATTAGTTTTTGAATCAAATCATCGTATTTATGTGGATCA 17663

Qy   829 gcattaaggtttgtaaattgagtcgacagcttagtcagttagtcaataatgtagcactttaatat 888
            ||| ||| || ||| || |||| || ||||||||| ||||||||||| |||| || |||
Db   17662 TTAATTAAGTAAGAAACGTCACCAAAAACCAAAGACATAAGTAAATTTAAGTAATTATAT 17603

Qy   889 ggcttctgtgctatatttttattttacatgctgtgtatctaggtttgcatttttagat 948
            |||| | ||| ||||| |||||||| ||||| || | |||||||||| | |||||||
Db   17602 TGAATACTTTCCCTTCTATATTTCATTAAAATAACATTTCTATGTTTTGAAGCTCATTT 17543

Qy   949 ttagtacctgtcagattaaaagaaacgaaagttaaattaaaaaa 996
            ||| || |||||||||| ||| |||| |||||| ||||||||||| |
Db   17542 GTAAAAATTTTATCGGGGTGAAAAAAAAATCAAACTCAAAATTACCAGAA 17495

RESULT   9

LOCUS   MISCVA   1200 bp   DNA   PLN   01-OCT-1996
MISCVA
DEFINITION   Yeast mitochondrial var1 gene (allele 40.0) a ribosomal protein.
ACCESSION   V00705 J01525
NID   g13610
KEYWORDS   ribosomal protein.
SOURCE   baker's yeast.
ORGANISM   Mitochondrion Saccharomyces cerevisiae
             Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
             Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
             Saccharomyces.
REFERENCE   1 (bases 1 to 1200)
AUTHORS   Hudspeth,M.E., Ainley,W.M., Shumard,D.S., Butow,R.A. and
             Grossman,L.I.
TITLE   Location and structure of the var1 gene on yeast mitochondrial DNA:
             nucleotide sequence of the 40.0 allele
JOURNAL   Cell 30 (2), 617-626 (1982)
MEDLINE   83050946
FEATURES   Location/Qualifiers
     source 1..1200
             /organism="Saccharomyces cerevisiae"
             /mitochondrion
             /db_xref="taxon:4932"
     CDS   10..1200
             /note="var1 reading frame"
             /codon_start=1
             /transl_table=3
             /db_xref="PID:g13611"
             /db_xref="SWISS-PROT:P02381"
             /translation="MKLKLLMILSMMNKTNNNNNIIINNTLDSLMNKKLLLKNMLLD
             MNNKKMNNMKRMLNNNMNPAGANPVVHRIGPAGNINNKIIMSKTINQHSLNKLNIKFYYNNDI
             NMEIMNTMNDKLINKLLYKMMTLKLNNMNINKIIMSKTINQHSLNKLNIKFYYNNDI
             NNNNNNNYYMNMNMKLMNIMNNMNNNNLCNILSYYNKKVTIEPIKLSYIYLNSDIF
             SKYISLNDMDKYNNGILTNYQRMLNNIIMPKLNDHNISMNYINNINNNKYNNMINL
             LNNNNNINNNNYNNNNNYIGNINNIYNNMTIDNIPMDILMYKYLVGWSIKFKGRLS
             NNNGRTSTTNLLNGTFNNKKYLWSNINNNYKLNYIPSNHNLYNNSNIKNGKYNIKVK
             LNFI"
BASE COUNT   605 a   57 c   67 g   471 t
ORIGIN

Query Match   4.78;   Score 46.8; DB 8; Length 1200;
Best Local Similarity   45.2%;   Pred. No. 4.2;
Matches 171; Conservative   0; Mismatches 207; Indels   0; Gaps   0;

Qy   203 atgaaacagttcaacagtaagtacttcaaaaacaaagcatcatgcgaagacagatgg 262
            ||| || || ||| | || ||| | || |||||| || |||| ||||| ||||||
Db   791 ATATTAATAATATTAATAATAATAAATAATAAAATAATTAAATTAATTAATAATAATAATA 850

Qy   263 agtacttgaataaccaacatggcgacaataacaacaataatgttacaagctcattac 322
            | | |||| || ||||| || || ||| || ||||||
Db   ...

Db   851 ATAATATTAATAATAATAATATTATATAATAATAATAATATTGGTAATATTA 910
            || ||||||| ||||||| |||| || ||||| || ||| || || |||||
Qy   323 gagataattatggaaatgagagatccattacttactggtggtggactattcctcttggcctccaagat 382

Db   17722 ATAATATTTATAATAATAATAACATTGATAATAATTCCTATAGATATTTTAATATAATAAT 970

Qy   383 cttatacatgtagctttgtaaaagggaattagatctgctcaagctcttggacaca 442
            || ||| || || ||||| ||| |||||| ||| | || || ||| ||| ||||
Db   971 ATTTAGTTGGTTGATCTATTAAAGTAGATTAAGTAAGTAATAATGGTAGAACTA 1030

Qy   443 tgaatgtttcatagaaggatagagccatttgagacaatcaccaccttagagatattaata 502
            ||| |||| |||| ||| ||| ||| ||| || || ||| || ||| ||| ||||| |||
Db   1031 GTACACTTAATTTATTATGGTACTTTTAATATAAAAAATATTATGAAGTAATATTA 1090

Qy   503 ggtattcctctcaacctaatctgaaccaacccaacctaacctttaccctagtcataacc 562
            |||||||||| |||||||| || ||| |||| || ||| ||| ||| || || ||||
Db   1091 ATAATAATTATAAATTAAATTATATCCCTTCTAATCATCAATTTATATAATAATTCTAATA 1150

Qy   563 ctagttttcaagaaaat 580
            ||||| ||||| |||||
Db   1151 TTAATAAAAATGGTAAAT 1168

RESULT   10

LOCUS   YSCMTVAR1   1200 bp   DNA   PLN   30-SEP-1988
YSCMTVAR1
DEFINITION   Yeast (S.cerevisiae) mitochondrial var1 gene, 40.0 allele.
ACCESSION   J01525
NID   g343948
KEYWORDS   repeat region; unidentified reading frame.
SOURCE   Yeast (S.cerevisiae) DNA.
ORGANISM   Mitochondrion Saccharomyces cerevisiae
             Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
             Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
             Saccharomyces.
REFERENCE   1 (bases 1 to 1200)
AUTHORS   Hudspeth,M.E.S., Ainley,W.M., Shumard,D.S., Butow,R.A. and
             Grossman,L.I
TITLE   Location and structure of the var1 gene one yeast mitochondrial
             DNA: Nucleotide sequence of the 40.0 allele.
JOURNAL   Cell 30, 617-626 (1982)
MEDLINE   83050946
COMMENT   Alleles of the var1 locus on yeast mitochondrial DNA specify the
             size of var1 ribosomal RNA. [1j; given sequence determines the
             smallest var1 protein. The coding region is highly unusual: 89.6%
             at with a 46 bp 'gc'-rich palindromic cluster that accounts for 38%
             of the total 'gc' residues. results strongly suggest that like
             mammalian mitochondria but unlike those from Neurospora, yeast
             mitochondria use 'aua' as a methionine codon. Comparison with
             sequence of a var1 allele specifying a larger protein suggests that
             some size polymorphism of var1 protein results from in-frame
             insertions of a variable number of 'aat' (asn) codons. [1] compares
             sequence with that of the 42.0 allele of ds401.
FEATURES   Location/Qualifiers
     source 1..1200
             /organism="Saccharomyces cerevisiae"
             /mitochondrion
             /db_xref="taxon:4932"
     CDS   10..1200
             /note="var1(40.0)"
             /codon_start=1
             /transl_table=3
             /db_xref="PID:g343949"
             /translation="MKLKLLNILSIINKTNNNNIIINNTLDSLINKKLLKNILLD
             INNKKINNIKRILNNNINPAGANPVVHRIGPAGNINNKLQHLNNINNXNTQIYNYNK
             NIEIINTINDKLINKLLYKIITLKLNNININKIIISKTINQHSLNKLNIKFYYNNDI
             NNNNNNNYINIINKLLYKMMTLKLNNINNINNLCNILSYYNKKVTIEPIKLSYIYLNSDIF
             SKYISLNDIDKYNNGILTNYQRILNNIMPKLNDHNISINYINNINNNKYNNIINL
             LNNNNINNNNYNNNNNYIGNINNIYNNITIDNIPIDILIYKYLVGXSIKFKGRLS
             NNNGRTSTTNLLNGTFNNKKYLKSNINNNYKLNYIPSNHNLYNNSNIKNGKYNIKVK
             LNFI"
BASE COUNT   605 a   57 c   67 g   471 t
ORIGIN   About 125 bp upstream of HinfI site.

Left column

```
Query Match
Best Local Similarity  4.7%;  Score 46.8;  DB 8;  Length 1200;
Conservative  45.2%;  Pred. No. 4.2;
Matches  171;  Conservative  0;  Mismatches  207;  Indels  0;  Gaps  0;

Qy  203  atagaaacagttcaacagtagtacttcaaaacaaaagcatcatggcaagacagatgg  262
                 |||  |     |          |       |     |    |
Db  791  ATATTAATAATATTATAATAATAATAATAATTATTAAATAATAATAATAATAATA  850

Qy  263  agtacttgaataacaacaatggcgacaataacaacaacaataatgttacaagctcattac  322
                 |  ||||   ||    |  |||||| ||   |
Db  851  ATAATATTAATAATAATAATAATAATAATAATAATAATAATATATTGGTAATATTA  910

Qy  323  gagataattatggaaatgaagatcattacttggtggactattcctttggcctccaagat  382
                 ||  |||    ||     |    |     |||
Db  911  ATAATATTTAATAATATTATAACTATTGATAATTCCTATAGATATTTAATATATAAAT  970

Qy  383  cttatacatgtagctttgtaaaagggaatttagatctgctcaagctcttggtggacaca  442
                 ||  |  |     |||||     ||
Db  971  ATTTAGTTGGTTGATCTATTAAATTTAAAGTAGATTAAGTAATAATAATGTAGAACTA  1030

Qy  443  tgaatgttcatagaagagatagagccatttgagacaatcaccacctagagatattaata  502
                 ||  |||  |   ||      |||||
Db  1031  GTACACTTAATTTATTAATTGGTACTTTTAATAATAAAAATATTTATGAAGTAATATTA  1090

Qy  503  ggtatctcttctaaaccttaatcttgaaccaaacctaactttacccctagtcataaac  562
                 ||  |||   |||    ||||
Db  1091  ATAATATTATAAATTAAATTAAATTATCCCTTCTAATCATAATTTATATAATTCTAATA  1150

Qy  563  ctagtttttcaagaaaat  580
                 ||  ||||
Db  1151  TTAATAAAATGGTAAAT  1168
```

RESULT 11
MISCVA
```
LOCUS       MISCVA       1200 bp    DNA             PLN       01-OCT-1996
DEFINITION  Yeast mitochondrial var1 gene (allele 40.0) a ribosomal protein.
ACCESSION   V00705 J01525
NID         g13610
KEYWORDS    ribosomal protein.
SOURCE      baker's yeast.
ORGANISM    Mitochondrion Saccharomyces cerevisiae
            Eukaryotae: mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes: Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1 (bases 1 to 1200)
AUTHORS     Hudspeth,M.E., Ainley,W.M., Shumard,D.S., Butow,R.A. and
            Grossman,L.I.
TITLE       Location and structure of the var1 gene on yeast mitochondrial DNA:
            nucleotide sequence of the 40.0 allele
JOURNAL     Cell 30 (2), 617-626 (1982)
MEDLINE     83050946
FEATURES    Location/Qualifiers
    source  1..1200
            /organism="Saccharomyces cerevisiae"
            /mitochondrion
            /db_xref="taxon:4932"
    CDS     10..1200
            /note="var1 reading frame"
            /codon_start=1
            /transl_table=3
            /db_xref="PID:g13611"
            /db_xref="SWISS-PROT:P02381"
            /translation="MKLKLLNMILSMMKTNNNNIIINNTLDSLMNKKLLLKNMLLD
            NMEIMNTMNDKLINKLLYKMMTLKLNNMNINKIIMSKTINQHSLNKLNIKFYYNNDI
            NNNNNNNYYMNMMNKLMNINMNNNMNNNLCNILSYIYNKKVTIEPIKLSYIYLNSDIF
            SKYISLNDMDKYNNGILTNYQRMLNNINMPKLNDHNISMYINNINNNKYNNMINL
            LNNNNNINNNYNNNNNYIGNINNIYNNMTIDNIPMDILMYKYLVGWSIKFPKGRLS
            NNNGRTSTTNLLNGTFNNKKYLWSNINNYKLNYIPSNHNLYNNSNINKNGKYNIKVK
            LNFI"
```

Right column

```
BASE COUNT   605 a   57 c   67 g   471 t
ORIGIN

Query Match
Best Local Similarity  4.7%;  Score 46.8;  DB 38;  Length 1200;
Conservative  45.2%;  Pred. No. 4.2;
Matches  171;  Conservative  0;  Mismatches  207;  Indels  0;  Gaps  0;

Qy  203  atagaaacagttcaacagtagtacttcaaaacaaaagcatcatggcaagacagatgg  262
                 |||  |     |          |       |     |    |
Db  791  ATATTAATAATATTATAATAATAATAATAATTATTAAATAATAATAATAATAATA  850

Qy  263  agtacttgaataacaacaatggcgacaataacaacaacaataatgttacaagctcattac  322
                 |  ||||   ||    |  |||||| ||   |
Db  851  ATAATATTAATAATAATAATAATAATAATAATAATAATAATATATTGGTAATATTA  910

Qy  323  gagataattatggaaatgaagatcattacttggtggactattcctttggcctccaagat  382
                 ||  |||    ||     |    |     |||
Db  911  ATAATATTTAATAATATTATAACTATTGATAATTCCTATAGATATTTAATATATAAAT  970

Qy  383  cttatacatgtagctttgtaaaagggaatttagatctgctcaagctcttggtggacaca  442
                 ||  |  |     |||||     ||
Db  971  ATTTAGTTGGTTGATCTATTAAATTTAAAGTAGATTAAGTAATAATAATGTAGAACTA  1030

Qy  443  tgaatgttcatagaagagatagagccatttgagacaatcaccacctagagatattaata  502
                 ||  |||  |   ||      |||||
Db  1031  GTACACTTAATTTATTAATTGGTACTTTTAATAATAAAAATATTTATGAAGTAATATTA  1090

Qy  503  ggtatctcttctaaaccttaatcttgaaccaaacctaactttacccctagtcataaac  562
                 ||  |||   |||    ||||
Db  1091  ATAATATTATAAATTAAATTAAATTATCCCTTCTAATCATAATTTATATAATTCTAATA  1150

Qy  563  ctagtttttcaagaaaat  580
                 ||  ||||
Db  1151  TTAATAAAATGGTAAAT  1168
```

RESULT 12
YSCMTVAR1
```
LOCUS       YSCMTVAR1    1200 bp    DNA             PLN       30-SEP-1988
DEFINITION  Yeast (S.cerevisiae) mitochondrial var1 gene, 40.0 allele.
ACCESSION   J01525
NID         g343948
KEYWORDS    repeat region; unidentified reading frame.
SOURCE      Yeast (S.cerevisiae) DNA.
ORGANISM    Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1 (bases 1 to 1200)
AUTHORS     Hudspeth,M.E.S., Ainley,W.M., Shumard,D.S., Butow,R.A. and
            Grossman,L.I.
TITLE       Location and structure of the var1 gene one yeast mitochondrial
            DNA: Nucleotide sequence of the 40.0 allele
JOURNAL     Cell 30, 617-626 (1982)
MEDLINE     83050946
COMMENT     Alleles of the var1 locus on yeast mitochondrial DNA specify the
            size of var1 ribosomal RNA. [1]: given sequence determines the
            smallest var1 protein. The coding region is highly unusual: 89.6%
            at with a 46 bp 'gc'-rich palindromic cluster that accounts for 38%
            of the total 'gc' residues. results strongly suggest that like
            mammalian mitochondria but unlike those from Neurospora, yeast
            mitochondria use 'aua' as a methionine codon. Comparison with
            sequence of a var1 allele specifying a larger protein suggests that
            some size polymorphism of var1 protein results from in-frame
            insertions of a variable number of 'aat' (asn) codons. [1] compares
            sequence with that of the 42.0 allele of ds401.
FEATURES    Location/Qualifiers
    source  1..1200
            /organism="Saccharomyces cerevisiae"
            /mitochondrion
            /db_xref="taxon:4932"
    CDS     10..1200
```

Department of Applied Physiology
National Institute of Agrobiological Resources
2-1-2 kannondai, Tsukuba
Ibaraki 305
Japan
Phone: 0298-38-8383
Fax: 0298-38-7417.

FEATURES   Location/Qualifiers
     source   1..1083
                     /organism="Petunia x hybrida"
                     /strain="Titan parade"
                     /db_xref="taxon:4102"
     CDS   109..933
                     /codon_start=1
                     /product="zinc-finger DNA binding protein"
                     /db_xref="PID:d1005618"
                     /db_xref="PID:g439489"
                     /translation="MALEALKSPTAAIPSLPPRYEDHVDMNNLDSWVGKSRSRPRIE
                     TPPSEEEYLALCLIMLARSGNGTTPGSDTIITTISKEPEKKNRELFPVHQETEQSY
                     KCSVCDKSFSSYQALGGHKASHRKITTTATTALLDDNNNNPTTSNSTSGNVVNNISAL
                     NPSGRSHVCSICHKAFPTGQALGGHKRRHYEGKLGGNNNHRDGGHSGGSVVTTSDGG
                     ASHTLRDFDLNMLPPSPELQLGLSIDCDLKSQIPIEQEVESPMPLKKPRLLFSMD"

BASE COUNT   328 a   216 c   198 g   341 t
ORIGIN

Query Match   4.6%;   Score 46.2;   DB 8;   Length 1083;
Best Local Similarity   69.2%;   Pred. No. 5.6;
Matches 63; Conservative   0; Mismatches 28; Indels   0; Gaps   0;

Qy 381 atcttatacatgtagcttgtgtaaaagggaattagatcctgccaagctccttggtggaca 440
         |||||| ||| |||||| | ||| || ||| | |||| ||||||||||||||||||
Db 408 ATCTTACAAGTGTAGCGTGTGTGACAAGTCTTTTAGTTCTTATCAAGCTCTTGGTGGACA 467

Qy 441 catgaatgttcatagaagatagagccatt 471
         | |||||||||||| ||| ||| |||
Db 468 CAAAGCAAGTCATAGGAAAATTACAACTATT 498

RESULT 14
PETZFDB2
LOCUS   PETZFDB2   1083 bp   mRNA   PLN   10-MAR-1995
DEFINITION   Petunia mRNA for zinc-finger DNA binding protein.
ACCESSION   D26084
NID   g439488
KEYWORDS   zinc-finger DNA binding protein.
SOURCE   Petunia hybrida (strain:Titan parade) cDNA to mRNA.
             Petunia x hybrida
ORGANISM   Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
             Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
             plants; Magnoliophyta; Magnoliopsida; Solananae; Solanales;
             Solanaceae; Petunia.
REFERENCE   1 (sites)
AUTHORS   Takatsuji,H., Nakamura,N. and Katsumoto,Y.
TITLE   A new family of zinc finger proteins in petunia: structure, DNA
             sequence recognition, and floral organ-specific expression
JOURNAL   Plant Cell 6 (7), 947-958 (1994)
MEDLINE   94348284
REFERENCE   2 (bases 1 to 1083)
AUTHORS   Takatsuji,H.
JOURNAL   Unpublished (1995)
COMMENT   Submitted (06-Dec-1993) to DDBJ by:
             Hiroshi Takatsuji
             Department of Applied Physiology
             National Institute of Agrobiological Resources
             2-1-2 kannondai, Tsukuba
             Ibaraki 305
             Japan
             Phone: 0298-38-8383
             Fax: 0298-38-7417.

FEATURES   Location/Qualifiers
     source   1..1083
                     /organism="Petunia x hybrida"

                     /note="var1(40.0)"
                     /codon_start=1
                     /transl_table=3
                     /db_xref="PID:g343949"
                     /translation="MKLKLNIILSIINKTNNNNNIINNTLDSLINKKLLKNILLD
                     INNKKINNIKRILNNNNINPAGANPVVHRIGPAGNIINNKLQHLNNINNXNTQIYNYNK
                     NIEIINTINDKLINKLLYKIITLKLNNININKIIISKTINQHSLNKLNIKFYYYNNDI
                     NNNNNNNYININIKLININNNINNNLCNILSYYNKKVTIEPIKLSYIYLNSDIF
                     SKYISLNDIDKYNNGLTNQRILNNIMPKLNDHNISINYINNINNNKYNNIINL
                     LNNNNINNNYNNNNNYNNNNIYGNINNIYNNITIDNIPIDILIYKYLVGXSIKFKGRLS
                     NNNGRTSTTNLLNGTFNNKKYLXSNINNNYKLNYIPSNHNLYNNSNINKNGKYNIKVK
                     LNFI"

BASE COUNT   605 a   57 c   67 g   471 t
ORIGIN   About 125 bp upstream of HinfI site.

Query Match   4.7%;   Score 46.8;   DB 38;   Length 1200;
Best Local Similarity   45.2%;   Pred. No. 4.2;
Matches 171; Conservative   0; Mismatches 207; Indels   0; Gaps   0;

Qy 203 atagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagacagatgg 262
         ||| || |||||||| | |||| | | | || | |||| || || || | ||
Db 791 ATATTAATAATTAATTAATAATAAATATAATATTAATTATTAAATAATAATAATA 850

Qy 263 agtacttgaataacaatggcgacaataacaacaataatgttacaagctcattac 322
         | ||||||||| || ||| || || || ||||| |||||||
Db 851 ATAATATTAATAATAATAATATAATAATATAATAATAATATTATTGTAARATTA 910

Qy 323 gagataattatggaaatgaagatcatttacttggtggactattcctcttggcctcccagat 382
         ||| |||| | |||| | | | || ||||| |
Db 911 ATAATATTAATAATAATAACTATTGATAATATTCCTATAGATATATTTAATATAATAAAT 970

Qy 383 cttatacatgtagcttttgtaaaagggaattagatctgccaagctccttggtggacaca 442
         ||||||||||||| ||| ||| ||||||||| |||| ||||| |||||||||||||
Db 971 ATTTAGTTGGTTGATCTATTATTAAAGGTAGATTAAAGTAATAATAATGGTAGAACTA 1030

Qy 443 tgaatgttcatagaagaggatagagccatttgacaatcaccaccagagatattaata 502
         | ||||||||||| ||||| || || | ||| || ||| || || |||||||
Db 1031 GTACACTTAATTTATTAATGGTACTTTTAATAATAAAAAATATTATGAAGTAATATTA 1090

Qy 503 ggtattcctctcaaacctaatcttgaaccaaaccctaacttttaccctagtcataacc 562
         ||| ||| | | |||| | || | ||| | || ||| | || |
Db 1091 ATAATATTAATAAATTAAATTATATATCCCTTCAATCATAATTTATATAATAATTCTAATA 1150

Qy 563 ctagttttcaagaaaat 580
         |||| ||||| |||
Db 1151 TTAATAAAAATGTAAAT 1168

RESULT 13
PETZFDB2
LOCUS   PETZFDB2   1083 bp   mRNA   PLN   10-MAR-1995
DEFINITION   Petunia mRNA for zinc-finger DNA binding protein.
ACCESSION   D26084
NID   g439488
KEYWORDS   zinc-finger DNA binding protein.
SOURCE   Petunia hybrida (strain:Titan parade) cDNA to mRNA.
             Petunia x hybrida
ORGANISM   Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
             Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
             plants; Magnoliophyta; Magnoliopsida; Solananae; Solanales;
             Solanaceae; Petunia.
REFERENCE   1 (sites)
AUTHORS   Takatsuji,H., Nakamura,N. and Katsumoto,Y.
TITLE   A new family of zinc finger proteins in petunia: structure, DNA
             sequence recognition, and floral organ-specific expression
JOURNAL   Plant Cell 6 (7), 947-958 (1994)
MEDLINE   94348284
REFERENCE   2 (bases 1 to 1083)
AUTHORS   Takatsuji,H.
JOURNAL   Unpublished (1995)
COMMENT   Submitted (06-Dec-1993) to DDBJ by:
             Hiroshi Takatsuji

```
     CDS
                     /strain="Titan parade"
                     /db_xref="taxon:4102"
                     109..933
                     /codon_start=1
                     /product="zinc-finger DNA binding protein"
                     /db_xref="PID:d1005618"
                     /db_xref="PID:g439489"
                     /translation="MALEALKSPTRAATPSLPPRYEDHVDMNNLDSWVGKRSKRPRIE
                     TPPSEEEYLACLIMLARSGNGTTPGSTDTTITTISKEPEKKNRELTPVHQETEQSY
                     KCSVCDKSFSSYQALGGHKASHRKITTIATTALLDDNNNNPTTSNSTSGNVVNNISAL
                     NPSGRSHVCSICHKAFPTGQALGGHKRRHYEGKLGGNNNHRDGGHSGSVVTTSDGG
                     ASTHTLRDFDLNMLPPSPELQLGLSIDCDLKSQIPIEQEVESPMPLKKPRLLFSMD"
BASE COUNT      328 a    216 c    198 g    341 t
ORIGIN

Query Match
Best Local Similarity   4.6%;   Score 46.2;   DB 38;   Length 1083;
Matches   63;   Conservative   69.2%;   Pred. No. 5.6;
                0;   Mismatches   28;   Indels   0;   Gaps   0;

Qy    381 atcttacatgtagctttgtaaaagggaattagatctgctcaagctcttggtggaca 440
          |||||| | |||||| ||| ||||||||||||||||||||||||
Db    408 ATCTTACAAGTGTAGCCGTGTGCACAAGTCTTTAGTTCTTATCAAGCTCTTGGTGACA 467

Qy    441 catgaatgttcatagaagagaggcatt 471
          ||   |||||||| |  |  ||| |||
Db    468 CAAAGCAAGTCATAGAGAAAATTACAACTATT 498

RESULT 15
MII2CG
LOCUS       MII2CG      58407 bp    DNA    circular    BCT    05-SEP-1996
DEFINITION  Methanococcus jannaschii large extra-chromosomal element, complete
            sequence.
ACCESSION   L77118
NID         g1500644
KEYWORDS
SOURCE       Methanococcus jannaschii.
  ORGANISM   Methanococcus jannaschii
             Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
             Methanococcus.
REFERENCE   1  (bases 1 to 58407)
  AUTHORS   Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
            Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
            Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F.,
            Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
            Merrick,J.M., Glodek,A., Scott,J.L., Geoghagen,N.S.M.,
            Weidman,J.F., Fuhrmann,J.L., Presley,E.A., Nguyen,D.,
            Utterback,T.R., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C.,
            Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.P., Borodovsky,M.,
            Klenk,H.-P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.
  TITLE     Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii
  JOURNAL   Science 273 (5278), 1058-1073 (1996)
  MEDLINE   96337999
REFERENCE   2  (bases 1 to 58407)
  AUTHORS   Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
            Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
            Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F.,
            Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
            Merrick,J.M., Glodek,A., Scott,J.L., Geoghagen,N.S.M.,
            Weidman,J.F., Fuhrmann,J.L., Presley,E.A., Nguyen,D.,
            Utterback,T.R., Kelley,J.M., Peterson,J.D., Sadow,P.W., Hanna,M.C.,
            Cotton,M.D., Hurst,M.A., Roberts,K.M., Kaine,B.P., Borodovsky,M.,
            Klenk,H.-P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-1996) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, Maryland 20850, USA
REFERENCE   3  (bases 1 to 58407)
  AUTHORS   White,O., Kerlavage,A.R., Adams,M.D. and Klenk,H.-P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-SEP-1996) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, Maryland 20850, USA
  REMARK    The starts of translation for some of the genes in the
            Methanococcus jannaschii extra-chromosomal element have been
            extended.
COMMENT     Address all correspondence to:
            J. Craig Venter
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD, 20850, USA

            Carol J. Bult is with the National Center for Geographic
            Information and Analysis, University of Maine, Orono, ME.

            Judith A. Blake is with the Jackson Laboratory, Bar Harbor,
            ME,04649-1500

            Hamilton O. Smith is with the Johns Hopkins University School of
            Medicine, Baltimore, MD, 21205.

            Gary J. Olson, Claudia I. Reich, Brian P. Kiaine, and Carl R. Woese
            are with Microbiology Department, University of Illinois,
            Champaign-Urbana, IL, 61801.

            Joseph M. Merrick is with the State University of New York,
            Department of Microbiology, Buffalo, NY, 14214.

            All other authors are with The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD, 20850.

            Predicted open reading frames were determined using GeneMark
            software, kindly supplied by Mark Borodovsky, Georgia Institute of
            Technology, Atlanta, GA, 30332.
            e-mail: mark@amber.gatech.edu

            Submission and annotation: Owen White
            e-mail: owhite@tigr.org

            Biological role information, putative identifications, sequence
            alignments, on-line name and sequence search capability are
            available at TIGR's World Wide Web site.
            (URL: http://www.tigr.org/).
FEATURES             Location/Qualifiers
     source          1..58407
                     /organism="Methanococcus jannaschii"
                     /db_xref="taxon:2190"
     gene            302..1051
                     /gene="MJECL01"
     CDS             302..1051
                     /gene="MJECL01"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL01"
                     /db_xref="PID:g1522644"
                     /translation="MMGEQLDYLLNIVKNDKKFLETLKNMKIDVDSDEELSKIFFIII
                     TFPFWLRCHDDIKNALRLIDKITNKGSFTKVLDNSMKEYHVFYIYDVDKLSDLIAEEL
                     KRLYLICKSEGKSYENIFFRLDRKSILKLFMKISPFKQLDVKLKNQLDFFEEYF
                     NNIEKIGYFSIRMENSRYNEHIKSLHPKLLISLRVDKVHIRLEAQIDYNINNENEEA
                     YKKIELMKLNLLTIGYKAVEKFLEEIYEEV"
     gene            complement(1082..1471)
                     /gene="MJECL02"
     CDS             complement(1082..1471)
                     /gene="MJECL02"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL02"
                     /db_xref="PID:g1522645"
                     /translation="MLMMSIILSMCIQMIRIKSGLSSVICVYSQMMNFRIFWDIMTA
                     FEELELPTIEAIVKLKKLLKLMESSKNSEVYYQGVIKRIIQRLSLLDQEKSLKENL
                     KFCVDNYILSQRTIEDLKTAIEIASKL"
     gene            complement(1374..1700)
                     /gene="MJECL03"
     CDS             complement(1374..1700)
```

```
                     /gene="MJECL03"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL03"
                     /db_xref="PID:g1522646"
                     /translation="MRDCKVQVGGFGEWIIKCDLELTFDEVMRGEIEELVEFIKSLES
                     LLKSSSEDSFRKNVLNALISTLKTLPWTSRLEYADDEHNIDVYPDDKNKKWIIKCNL
                     CLFPDE"
gene             1934..3253
                     /gene="MJECL04"
CDS              1934..3253
                     /gene="MJECL04"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL04"
                     /db_xref="PID:g1500652"
                     /translation="MFVDREEELKALNEKLDSNNFEFIVIYGRRRIGKTKLALKSVEN
                     REHIYYLAVEGDNLKHFKRYASKVEPTIEYAKEDWEAYFNFLKDKIIIIDEFPNLIKE
                     NPNVLSLFQRIVDIHLKNTKTKLIIILGSSISMGERVLSYKSPLYGRKTGVKLIKPLK
                     FKHLKEFFPKAIWEELVEITGFADGIPYYLEKVDLDKEIKRVDSFLRYEVDF
                     LMKYEFEEPTTYKKILEAIAFGNHTLGEIKNYLGFKHSDLTPYLKNLIEVFIERQTP
                     ITESVKSKKGRYYIKDNFIAFYFRYIFPNLSAIEEGIFDIEEIKADYNQYLGFVFEKV
                     AKEFLIELNKMNKLPFKFLKIGRWWHRGEEIDLIALNDNDKKALFVEVKWKDLKRDV
                     KKIYRDLYRKSKLVGLDDYEKYYAIVGKKIESKENGDCLLFDLEDFS"
gene             3265..3453
                     /gene="MJECL05"
CDS              3265..3453
                     /gene="MJECL05"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL05"
                     /db_xref="PID:g1500645"
                     /translation="MEIIALLIEEGIIIIDKKVAERFLKDLESSOGMDWKEIRERAE
                     RAKKQLEEGIEWAKKTKL"
gene             complement(3784..4134)
                     /gene="MJECL06"
CDS              complement(3784..4134)
                     /gene="MJECL06"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL06"
                     /db_xref="PID:g1500653"
                     /translation="MSVELRTLFRLIAVLEHSEEFKVLFACERHFESGYCKCGPMEM
                     CNIALAEAMKEDPTLVLRKWRRVFTYLEEVGIIKTRKLEAPANRPRRYIKLSENRMEA
                     LRTAIDKEYEKLIR"
gene             complement(4558..5805)
                     /gene="MJECL07"
CDS              complement(4558..5805)
                     /gene="MJECL07"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL07"
                     /db_xref="PID:g1522647"
                     /translation="MIKMFIISKNHLDIIYNFVDWEFKNISSIGNIINDEVVWNELPK
                     GVEGVLCGVDGSRGKVEFCSGIVYGLSSYAIGKNIEKGMFELGVLPFFKEEDRVRRLM
                     MTLEYRLATLVSKNVDLILLDGTLSGALIMPPLLSGDTNPLTVPPDLAEDLGWKFIKS
                     LDNFWDEVLENLDGNIYDNTLLAIKIFQKFDSTYSEYVEDIREELFAANILNSRLEIA
                     CWGVFYEYIELLHSLNRLLEYDCTFIAKNFENSIITEKLKENNINVDILLDATLLNQI
                     FRGRGYTTLKLEDCYNKKRSRRHINKICDVFGKFKFLEVIRENPEMIPKTYVRFAE
                     SSPILALEVPRTNKKSIEEVISLLIPYSKLGYPRYLKDAHNKAKISKKEFKKQILFMI
                     KYISEKNRDFSLFFQSGREVLGE"
gene             complement(5829..7391)
                     /gene="MJECL08"
CDS              complement(5829..7391)
                     /gene="MJECL08"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="M. jannaschii predicted coding region MJECL08"
                     /db_xref="PID:g1522648"
                     /translation="MMVINSVVVGTVVASKNVNEFEFVIENQVIDKIKKGEFVITKNT
                     HGDYLLSKITKYSVSVNALIGDKSEDASELAKIRGVIYSEEMLNNSSKFLASAKILGVI
                     NNESGSIESNVYPINVPQNVLTKDDLLAKIFSNGSIEVGLVKVRSSTKVKLNAKELC
                     SRHFAVLAMTGAGKSNTIAVLVQELFEKDKGKMNIVIVDPHGEYVKMRNTHILPAKLN
                     PILVPEHLAKLLGIGENSSVQKSFLVYAALTVKYECKENKQISGLEYLKKIEEKLV
                     ECADKIANSEDKKRIYIEYYDGTRCRXKTVKKDDEMSINRVIEKLRWFINKNKNILGE
                     NEGMFDIKSDKINVLPLQKIEDEEAITIVGEFLKRVLKERIKSVHDEIVEIKALEKPT
                     LVIIEEAHLFAAKNLKDRSGYWINRIAKEGRKFGVGGLLVSORPKELNPTVLSQMNTK
                     IILRIVEPTDQKYILESSENVGEDLLQDLPQLSTGEAIVGSSLPLPALVIKKYDGV
                     LGGEDGLKNLKI"
gene             7780..8106
                     /gene="MJECL09"
CDS              7780..8106
                     /gene="MJECL09"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL09"
                     /db_xref="PID:g1522649"
                     /translation="MNYMKFMWNSEDGFIVEQHKEIPFEDWIVNTVEQLKNLQYTR
                     YTFEELTPKTKNLEAYNKLLEILKHTNMDIENNVKRIYHCIAENNNYKYEIAFYIVI
                     SKNWNE"
gene             8128..8787
                     /gene="MJECL10"
CDS              8128..8787
                     /gene="MJECL10"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL10"
                     /db_xref="PID:g1522650"
                     /translation="MHIHLFRCQCMIETIHIKNFRGIRELKLENLGQINIIAGKNNAS
                     KSSILEALALFLSAKEGFSLFIKILREILLWRGWYGEKSIYDLFYKNSKELEVSVKFL
                     NQDFANTLKNSNQSFANKNIAVELKSDKNSWSGRFDSHLIHPDYISSILTSAEATQS
                     NFEFITSLTLIKFGYIESIYSQAYETQVLQDAIRLLREAYPEVKSLSPLQKYNKWIIH
                     V"
gene             8818..9162
                     /gene="MJECL11"
CDS              8818..9162
                     /gene="MJECL11"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL11"
                     /db_xref="PID:g1522651"
                     /translation="MGEGFKSALIIALLSILKNGYLLIDSAEAFHPSSLEITSOML
                     TKSVKNNNVQVFLTTHSLELIDFLLEHASKEGIEGRLIYMRRDGENLISSMESFENVR
                     EMRETLGIDLRG"
gene             9165..9890
                     /gene="MJECL12"
CDS              9165..9890
                     /gene="MJECL12"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL12"
                     /db_xref="PID:g1522652"
                     /translation="MRILLLEGITDVAFFIPLKKLYGFSEISCDGIIRAEKMGDISK
                     PICLENEDVKLLVFHSGGKSKQKHALTAMLTAIKMGVLSNIKLGIRARDIDEDHDVKN
                     WTKSIIKNAGFEVKEGDKFLIIEDLNKIXIAVLGIANTDEDDFNIPSFELKRELEAVIT
                     DMAKEISIIEFKNSLESLSNDAERRLKPKDITHVLAIAKNFDGDSMSGLYRKFIEEQ
                     INNKNKVNFLLLLICILPCLTIF"
gene             10678..12486
                     /gene="MJECL13"
CDS              10678..12486
                     /gene="MJECL13"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL13"
                     /db_xref="PID:g1522653"
```

```
                     /translation="MINMCYNFENVKKAYEDQLNYIKNSLIPNLSSHANNEVKISM
KKLANLGFGVLVDSTLSSSITGYYKVRDWIELLKELQKSGENKCCNIVLCDYPEFE
ISMREIPYYVNKVVKFNGVIISASYPCVLSKKHLYICPKCGRIKEVYSELFWDDKVF
CEFCCGKMEFANVMDYENFQELVVQDLSDESEYYGIEKNPIVWYCGAKPYFGHVKIT
GIVREVPRSSKSRIYELIVAQAINVEKLGVEKSLINLTEEDVKNIKKVAKRGDIIDILA
DILIPPLLCDDAIVRKAILIQQIAPYLEDIGKINLLVTEVGIDKTAILKRIGNIPGN
NFINIAALKEEELATPYDKRSNIIGFKFYTVCGGVIPRTLGVLCIDDFNENNKLSTKLS
EAFERNVLTTNKGSFYCVPAECSFLCACYPKTKFRKFDQKKSIIKQIGISSILLKNFD
LIFPIRDIPDKDRDEVAKYIFLKYINSDNEEIEGYDYFVDVGGEKIKIDFEFLKKY
VVYSRQITPKITDEVIEKISNWYDEMRKNHYITFAKQLNTVIKLSIAVARALKECVDE
DDVKEAIDIIMHYLKQVVYNPKKGIDVILLYKNKT"
     gene            14486..15430
                     /gene="MJECL14"
     CDS             14486..15430
                     /gene="MJECL14"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL14"
                     /db_xref="PID:g1522654"
                     /translation="MLSFMLISESMIFTSMDNFVEALFEIDENSKEKKIKSYIESFTK
GVNDIINLYIGIKIPEPILDKFFGEKEKGGLVFKFIRDLFMSLNKKGVQPVFILDELQ
MIKDVMNGGKPLLKSLLFQEVLSLTKERHIAHVFCLSSDSLFIEYVNAGELEGRAKY
LLVDDFDKETSLKEMDFLAVEGNINLTNEDKELIYSYVGGKAKDIKYVIEESKFKDLR
EVLDFMLKDEVSKLRKLLVKIKTKKIAEVEYENVVKALLFKDNYEINEYLMDENTKE
FLIKRNILFLNPVEEILKPQSYLVWNAIKRVLQNINDF"
     gene            15447..16544
                     /gene="MJECL15"
     CDS             15447..16544
                     /gene="MJECL15"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL15"
                     /db_xref="PID:g1522654"
                     /translation="MKFFNREKEINEILILEEEPNNIYFIYGPLNSGKSTLIREVIT
NRLDKSKYIPFFIDFRTRNILNVDNFIECLFEVDEKSKIDDFREYAKSLADLLVKGSE
EISKYLGMFIKVPKPFFIDFRSRKDKSADVYQYIEYLFAKLNEKGKKPLIFDELQM
IREITLNGNRLLLWSLFQELVALTKVQHLCHVFCLSSDSLFIEYIYGKAELKGGVDYI
LVDDFDKKTALKFMDFLAKQKNINJTNEDKELIYSYVGGKAKYIYDVIVKLKAVKDLK
YILETKLEEERNHLEELLEELKEVEEDYEGINYDEVLEALKLFKDNYELPKSKIKRRIRIF
LIKENILFLNPQKGTLKPQSYLVWNAIKRML"
     gene            16602..16814
                     /gene="MJECL16"
     CDS             16602..16814
                     /gene="MJECL16"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL16"
                     /db_xref="PID:g1500646"
                     /translation="MSILISNKQFNHGLKDEFATKKDLELLEERILRYVDNKFNQLDK
KIDRTFYLLVFFIILWVSREAFFYLI"
     misc_feature    complement(19889..20104)
                     /note="MJECL17: similar to GP:U08838_1 percent identity:
58.7: identified by sequence similarity to archeal
histone; no plausible start codon"
     repeat_region   20485..20749
                     /rpt_family="IS_C1"
     gene            20873..21508
                     /gene="MJECL18"
     CDS             20873..21508
                     /gene="MJECL18"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL18"
                     /db_xref="PID:g1522656"
                     /translation="MIFMFHSGAMELLKIAEKLYDKDSEKAVEVYDKAIKKAERIYDD
YAKAVILSNIAKSLYSRGLTNKVIEVYNAIKIAEESSKRDVILSKIIENLCSNRLID
KALEVVNISDDSKAIALSEIAKAQYNIGMHDEALKNYDKAIFITEGVFDEIKSSI
LFEISRDFYNYGLVDKAFEVIGKIPYSKYRFRLLDKMAEDLHKNIKYIHGE"
     gene            21498..22022
```

```
     gene            21498..22022
                     /gene="MJECL19"
     CDS             21498..22022
                     /gene="MJECL19"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL19"
                     /db_xref="PID:g1522657"
                     /translation="MESKEFDDDKIKIISMTPKSEKSDKIVMQIVEEFINSFPDDRYK
FRVLLKVAELICKNGLCNEAFLILDKIPDSYYKSSALYKMADILYRNKEHDRLIQIAE
KIPDDYKKSEVLLKVVELLCESGKYDEAINAEKIPDNYYKSEALFKIAETLSNKGYY
DKAVEIAEKIPDNF"
     gene            22850..23293
                     /gene="MJECL20"
     CDS             22850..23293
                     /gene="MJECL20"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL20"
                     /db_xref="PID:g1522658"
                     /translation="MGKCRHNGEVSIFGVRPASFPYFPFNLMDRIGGFVILDELWLRR
WCEIEYPMKIPTLVPIEDYGIFEYMDLIVDFIKYHVSNGREVVSCIGGHGRTG
TVLAIWAGLNGVENPIEYVRECYCECAVETEEQEEFVMEYLKKRL"
     gene            complement(23295..24593)
                     /gene="MJECL21"
     CDS             complement(23295..24593)
                     /gene="MJECL21"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL21"
                     /db_xref="PID:g1522659"
                     /translation="MWRGVMVVQQFINRIRNCPLARDIIKKFFIPVIKERKNIEKEV
EDILKVLRKNYCSGEYANKDCICAWFKFFEDVGRLHEALELKAFWEIIFPNHLNDLSE
SLQLFISNYAYMRPFGSYSSYSIRVRIHGFIGAVCIDRNFNLHNNTWISNICNCYENNL
RMLGLGFYTSDTLDPRYNINPSCNINSTTTNKCGLLYIISNTQITASLVSYLLNLLRN
GNIYDAYNFLRIRGVGKPKIACLFLRDISIIHNLMLYGDCRDLLYKPIDRVKKVIEE
LVKCKCKLNYNFDINEVLEKFDYTECRCIQKLRKNAKIDSNDFKTLPYKIFLYEYSK
CCKLDQRYVEMGMWFFNSKVAKYINRRLCKYTIQEACKDLIKRLRKAINNTNIRYIGN
NQFEININGKIIVSEDELIDLILKGEIEIYDELDIIFF"
     gene            complement(24851..25090)
                     /gene="MJECL22"
     CDS             complement(24851..25090)
                     /gene="MJECL22"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL22"
                     /db_xref="PID:g1500667"
                     /translation="MLELRSLSIRGGIAVLECGRCISAINLKTGDKIWEFKTEWDIE
SISIKDNRVMLKCNGRRHYIDLKTGRKIRELIIL"
     gene            complement(25133..27616)
                     /gene="MJECL23"
     CDS             complement(25133..27616)
                     /gene="MJECL23"
                     /note="identified by GeneMark; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJECL23"
                     /db_xref="PID:g1522660"
                     /translation="MVRMVFVENKINGVENIRQEIDNLNVSNYKEIYDKFKNVFSKN
RDIFNYYVDKLIELMKNLDDNELSLEIDKFMLCILKDSIIWEFKAGGSVWDLSIKDIK
DNIILGCSNLFALDIKTGNKIWEYKVEHNVDSLFIKDNIVMLEYRGGHVCVLDVMT
GDKIWESKVGERMWGFSLKDNIVILGDGKDIYAIDVRTGKLWEFEEAEWCWELSIK
DDKIILRCEDEYGCEYFYVLDIKTGEKILEFGGEWHVSDLLISGDVTILADMWGCVYA
LDTNLYSKIQRVRPQLTNIMKEIVKIDLTLLKKSLNLNEWDELPIQITNKSLKDITIS
KISIINEEDLFKDIEPIKIRGRDFKVINLFINPKVKGKLPIDIVVEFEDFINRYKE
RFTEVLTIIKFKGDNVDDMRPEKIDKILKQEIDNLNTSNYKEIYSRFKNIFNENRAVF
NYYMNKLIELVNNSNDELAINVGKFPLDIIGIKRVDELLMEFRAEGGVRLLSIKGDIV
ILGCVSGHVSAIDIKTGKKIWEFKAEDTVWGLSIKDIVVLGCGNIFESIVMLKNGKI
LEEGYAYALDINTGREIWRSKIKHDVRSLSIKDDIVVLGCKKGYILALDINAGNMLWE
FKAKSGKSIRNLSIKNDILLFGCDNYLYALDIDTGRELWRFKAEGEVKSLSIKKDNVL
```

```
                      LGCRGGYVVLLDINTGEKMERFKVVGSVLRLSIKDDIVILGCNRECVYALDINAGENL
                      WAFKTDGDVNGLSIKNDAVLLGCDNYLYALDINTGEEIWKFKTESAVLDLSIKDNIVI
                      SGCKRGHVYALDFNIIKNYSIIQKIKQVL"
     gene             28069. .28848
                      /gene="MJECL24"
     CDS              28069. .28848
                      /gene="MJECL24"
                      /note="similar to SP:P37522 percent identity: 35.7;
                      identified by sequence similarity; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="SOJ protein"
                      /db_xref="PID:g1522661"
                      /translation="MVVISIANQKGGVGKTTIALNLSFTLAEKGYDTLVIDLDPQFNL
                      SFGIIGMKLLDYADKNIGILLSKNSVKKKEIEESIIKINDKLDLIPSHLQLSAVEKML
                      VNAYAREMKLKNIINQIKENDYIIIDNAPSLGLFLINSLVASDYIIIPCEPSYFSIA
                      GVOLMLDTVEEIKESNLNPKLKVLGFIFNKYSKOSKIPQKRLEQLKQLYPNIPVIGVI
                      PRTITVEKAEREGKPVFKFDANNPASVAFSELAEWIENVK"
     gene             28838. .29170
                      /gene="MJECL25"
     CDS              28838. .29170
                      /gene="MJECL25"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL25"
                      /db_xref="PID:g1522662"
                      /translation="MSNDDNALKLKEISSGTVKQITSLPKKEPQEEKMSKTLRIKN
                      TTHEKIIEMYGKKVGSQGEVVDKGVAVLVALWKILPEEQFKRVVKLAEEDRFEEFADR
                      LGIEIKEE"
     gene             29175. .30206
                      /gene="MJECL26"
     CDS              29175. .30206
                      /gene="MJECL26"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL26"
                      /db_xref="PID:g1522663"
                      /translation="MGEIMFYDREKELNYLKTYCQLEPNSILFYGPKSSGKSTVMR
                      RVIKELENSNIVFFYYNLRKYATPTRDEFLRVFFEKSDKKYLLNKLELNLGVCKFGIE
                      ENFDFNNLCLNDVFAKINESINAVVEEGKKPVLIIDELQKLKNIYFNGGKSLLNELFN
                      LFVSLTKMEHLCHVCCLTSDTLFIDEIYRNSTLKNASEYYLIDWLRKGTIRNILKEEG
                      FSEEEINYALDYLSLPYEISQLINNKKLGLSVEQTIKQWINVEKDGLKYLIDTTDLDE
                      EGLINVLSKFKDKIKISYDREVKKEEMKYLKFLIENEILFYDVINGIIKPTSIIEWHA
                      IKEININAMQ"
     gene             complement(30558. .31056)
                      /gene="MJECL27"
     CDS              complement(30558. .31056)
                      /gene="MJECL27"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL27"
                      /db_xref="PID:g1522664"
                      /translation="MDILGEIVNVDEYVEKLELQKNDIGFYKCPFCDYTNADAKVRK
                      HVKSKHLEEIEKELKKLESQKSKNNGKKQTGQKKQGKGKKQPKRVRETCVSTOERKDY
                      VLFFCHNHKVRLHLANGEVLEGKCCCKDPYTVLVDVGKGDVVINKAYIVKYVPLDLE
                      KL"
     gene             complement(31531. .35349)
                      /gene="MJECL28"
     CDS              complement(31531. .35349)
                      /gene="MJECL28"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL28"
                      /db_xref="PID:g1522665"
                      /translation="MNNILSNSIKSLRYWDDQTVKAIFDKYQINVSGIYNISDDILES
                      DLKLHIAQLFLIVKHKEFSVLGETHLEKLIKLKLVKQFKYLEIRFLGKFKGKKQFF
                      IDISDLHVDNILEKTEEIIKLIKKQYEYIKKQKKRLDGYVGIAPRKEKRGDAGAAT
                      AWSYIAFDFDVEEWKTNKMPTEEEIMEKLIKYLSKFVEKDILPHKVAFTGGGLRFILY
                      PERPLEEELILLRMIAEDLGADIAMYDLARVDRLVGTYNYKEKYGSPRPCVTIAQLT"


                      DEEDIEKLLTPIILYKFGIIENYKEFEVLFNEKREELYNKRGLNNITTTINRLIKKP
                      NLNKTILQDINKRIYKWLYTIQDKLTKKLGKRWIEKLLSVLGIDYKYGKNGTRLDLWS
                      LFFDGKNPDCSIYINEGYNAVMVDFHDPNMRFIAIAGLWMIKEFRDKIIEFLKLHNI
                      NPKPSYRTVREIMNELLDRETIKIEAEGYLPKEAIIKAYOLSIEQGVPVFLKADFGR
                      GKTYLTRNTREIKEVFKKHAIAVAFPYKIOVLQVGAGLHADGVVPMYEDGKKLDK
                      NTIHYLTIGTYDQVENMLNDLCYDYKGERIQVANEEDILLAIDEAHDLVIQKEFRKR
                      AITGVKKCIDRAGGCVLLTATPELINLNYPIEVEFKDEKKLFERCSIHIAKNIGE
                      FCEYILLMFRQGWIKNAVVLVDNKMIENIKYTLELYGFNKPIYVTREVGIDKASK
                      MIINEKVPEGLLIATRVISEGVNIKNHVDLVWALYCKSATTIRQFIARCRNGGGEL
                      IVTAPFKEREEPMIDYNAMIEMFKENYKLLKEYLETDIEVLKELDKNYKKVLISQI
                      QNAIYDEEKKDWVLDEDEIAHIYNSLLEAHITRDYKLLKEYLEKTTGYEFAIKTIKE
                      LKESKLDKFLKYNYLEYLEKVSAKHIIIAFKEYGVNEIKNALEYNNYKKKFDCRDEYT
                      PQLIKKHSKRINRSINVLKDVYNIPIEVYNLMLNKGKDDDKKSNLDENTIKKCLEE
                      INEYKKIVKEINENDRKTKKIYHIDLKLLEEHKDKLSTLARIIEELILKTFLCPPSK
                      WGKIQRIIRAVWNIVVGEDDERLDRKLGLRSLVAKVLIKVRDFAIERQKFKIRELIE
                      VIKEECGFTLDEVRRLIRAIFNCVIRGVKKLGENAVVEIKELNKEFKTLYEIIKQN
                      INSNSVEKCEKVIEKKIEENGGEILEMELYELIVEKLKFVEEVFYKALEKLKKLGVIY
                      EPKPGGLLRIT"
     gene             complement(36217. .36420)
                      /gene="MJECL29"
     CDS              complement(36217. .36420)
                      /gene="MJECL29"
                      /note="similar to GP:U08838_1 percent identity: 64.2;
                      identified by sequence similarity; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="archaeal histone"
                      /db_xref="PID:g1500647"
                      /translation="MTELPVAPFERILKKVGAERVSRAAAEYLAEAFEEIALEIAKEA
                      VDLAKHAKRKTVKVEDIKLALKK"
     gene             complement(37148. .37609)
                      /gene="MJECL30"
     CDS              complement(37148. .37609)
                      /gene="MJECL30"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL30"
                      /db_xref="PID:g1522666"
                      /translation="MSPSNVFIDSSVMGLFIGDEKAHLSKLINDGFMLCINPIVF
                      SETMFKVTFHIALEDGIRGVYDLKKNLNRYSWVYEEVREKIDKMIKMNYLKILDTNWE
                      VLKLAPEIGKKYNLITNDAIIIATCKYYRINKLAITFDSDEFKVDFIEIIKE"
     gene             complement(37596. .37802)
                      /gene="MJECL31"
     CDS              complement(37596. .37802)
                      /gene="MJECL31"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL31"
                      /db_xref="PID:g1522643"
                      /translation="MIYMSEIIEVIYEDGVLPLKPLKIKGKKRLKIVNDDVEEFL
                      KSMIIKCKDIDYKKLKEAYYESF"
     gene             complement(38825. .40141)
                      /gene="MJECL32"
     CDS              complement(38825. .40141)
                      /gene="MJECL32"
                      /note="identified by GeneMark; putative"
                      /codon_start=1
                      /transl_table=11
                      /product="M. jannaschii predicted coding region MJECL32"
                      /db_xref="PID:g1522667"
                      /translation="MGKELDNRTLDFYFEKYGKKKTSTEPSYKDIIEYITALISENLE
                      DYQSQHTKLINIILNLVYPIKYFPHHLYDLGYRPKKGSTLGVRIVLDGIYIYTKNGNPA
                      EITPDVILVNNDTHVLIFECKSKSIKKEQLEKYLIKENLHVINKGYISVTKNPRL
                      YKSDVSYMSFDDLTTHPVLKDTEDIQILKVSPGEISLERGNYENQKLNKIFPITWGEK
                      EKPSYDLLICDAENDXEGSLFKILILRELVSMALSGHDEPEVGMVFSKKIILPKGEYS
                      LESLKEDDFKEIENIQKLCKKCIIDKENKLLRVDKDIVLPENKPLNANVLKLIDA
                      LNISINMDSLVEKEFTVGTIMDALYRSTSISIIDYFYDYKNIVSKRIEETLKLFSNR
                      NLKGYLKKVSRKWRIKVIKDKRMARKFKESCKEEIGAIKSWQTTLD"
     gene             complement(40122. .41348)
                      /gene="MJECL33"
     CDS              complement(40122. .41348)
```

```
              /gene="MJECL33"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL33"
              /db_xref="PID:g1522668"
              /translation="MIIMDTIFVGSAIPDITNRPOLYEWLERNMERSYKDILDDTKLE
              PEQNVVKTYILESNIHPKKLSRIKTGKIKPLDEYEFYMLRIGDEGYFFIDAKNPRFW
              KLYTLQKSEESDKYFNKLISPIKSRLDNLWMPTGDLEKFLKRADYVRGLSTKHDETPF
              YLDNDVEERANKLSISSNGESVVELIKIIQSLSSDEIQEFEKLIKDFQLLNDTKIKKI
              FNTIKKLHEFKHLMRITKSKIKIVSEDDKDKFVLEDIIYWGKFTVKGYSIEKKNEIVD
              KTIENYEQKIEIEDSLIDYTSSDWGDRIPLIYEFEKEIEDLKNFVEGLISVKEPFKI
              WGLAKQVEDMYYISGVDLHNGDRFSLEVTPWWMRLYLPKGSCGNTALRLLSNIQQTY
              DSETILEVEKYGERIR"
gene      complement(42228..43115)
              /gene="MJECL34"
CDS       complement(42228..43115)
              /gene="MJECL34"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL34"
              /db_xref="PID:g1500677"
              /translation="MKHMTPGLTEKKLVYYDERYPYSWIPHNVAKHISKELEKYDFE
              VIDAPSLQRVLGEGSKDTNKEYVVFAQDVVPDLVLDNPDNPTSNSLLRRFLNAGQI
              VWMGDSPQYVGRADDNQKLNPPVMQNVLSMNPNVITTNKRISLTISGIVLSLPLWVGT
              KPHNKTPHGVSGVSINPLAVAIDNVQYAHAFIISYKQSASGFVRIYDFPINKEELVT
              KQFIRGILGVATRDVTTPIWNKIEILSEEVNKIDEKFDTKFNTLKSEIDTFKTTINEI
              LKLEKEILEVIQKKCENKE"
gene      complement(43112..45004)
              /gene="MJECL35"
CDS       complement(43112..45004)
              /gene="MJECL35"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL35"
              /db_xref="PID:g1522669"
              /translation="MKIEQVKIENFRSIYDLTVEFGKVTVLIGKNSSGKTSILNVIRI
              IFENLDTNSTERNLNITITEQNLKDQIRRLWFYTYQNYPIKIAVWLNFDEGEGDKLR
              EVCGLESNVKKAYIEVSIESNSNONIIWKLQNVIIFGYVSETSKIAKEALKKMKEIHSK
              VTGHNIESKTGDSLVPCVIVESGTIKNKDVFNIFIDLIKNKICYVNTFPEIYKMVNIP
              IKGLFAHPMVSSELINVKNTIKDFARLRMLYDCLREIKGSVDTYEQPDRTEILTKSY
              EGMSLNYELFGSGDQIIDGILASIISKGKHHFLIEEPEIHLDPAYIRLARVLEDLA
              NNLNVQLIIVTHSPDFIATLKDKRFVIGVRKDSIKINFEGYPTELLATRIFYPFRGSR
              LSESLARDLGVVPGYFLFSDVAILVEGNSDRIILQGYIDLLIEQLKNLPKFSYIIPY
              AQRSLRKMKVEQPYFTKFLIADNDEGLKNVKIAKDMGLRENFEVFIIEKKDILG
              YIKPEDFILADELLRECLKDKYEEVLQDEEIKSIIDEIKIFGACKNEEGKNEKDLLH
              KLSGKLFFVLKNVDEEFIITLGWDSGRSIERYLKCKIAEKIVSRLTEVPEDIKKIILN
              IDSELG"
gene      complement(45391..45834)
              /gene="MJECL36"
CDS       complement(45391..45834)
              /gene="MJECL36"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL36"
              /db_xref="PID:g1500679"
              /translation="MDLMAYWLCITNEDNWKVIKEKKIWGVAERYKNTINKVKVGDKL
              IIYEIQRSGKDVKPPYIRGVYEVVSEVYKDSSKIFKPTPRNPNEKFPYRVKLKEIKVF
              EPINKFKELIPKLKFITNKKRWSGHLMGKAMREIPEEDYKLIVGN"
gene      46155..46868
              /gene="MJECL37"
CDS       46155..46868
              /gene="MJECL37"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL37"
              /db_xref="PID:g1522670"
              /translation="MNNFSKNLEECFQKYGVENFTLLAGYIAKYPKYEEISNNTKFL
              NEYHEVSNITTCLHVYKNLAEGVKRDFNLDLLYLNDYEIKKVKSMVEPNVISDAITV


              VTLVNDYNNLIDAARNVKKGDKESYTKFYIALGIVFDVILIKENVAYKVSYKLVGIL
              ISKTGFYKVIYKYGGSTALKPIESCTHWISRGEINSMPSKLYNNSDKLTNISIKINTS
              KLYNKSINTIRSKIMEVSS"
gene      complement(47194..47973)
              /gene="MJECL38"
CDS       complement(47194..47973)
              /gene="MJECL38"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL38"
              /db_xref="PID:g1522671"
              /translation="MANMQSLTNIEVQRFHDCEWEYFKEFDDEFNKLWNEIEKTLGRD
              FINYLSAYFQKNLVYMLGKEFKLKLVVDTNIIFSQVLSVTKGELPWILDFINNPFIE
              LYAPQLIVDELKNKIENVLPKKCKKKKNIDENKAKSKAIKIANILLSNIKIINDKKSNN
              WSKIAYNLIGHRDVKDIPFVTLALSLDTHGIITRDKDFKDQKIIKIWKVGEVKVVLTE
              LSQSFSFCIMNVTAPLAFKICTSIIITLEIVTSIIKKTN"
gene      complement(48326..49381)
              /gene="MJECL39"
CDS       complement(48326..49381)
              /gene="MJECL39"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL39"
              /db_xref="PID:g1522672"
              /translation="MPILLEHIQLNDEDLKDTKHLAEILKYKNNSIKTIINVLEWEDS
              NIRYCYEKSNVYFITFFIIVGLVWAIFPPEVWLWCEQVFCISPTIHIICCLYFIITI
              ILFLFLCGVVGTFLHLWATFFTLSKCDSKILKLKEGLFTTLSLIWISTSLKTILKTKY
              AICRDYAKLTSAILHNLNIKHYFLVYPTHVAVAVKIDDYYVIDQKLPYKIDVWLKK
              LGKEKVKIYTPVDIYNSKLKFVEKYYKNENNLKSEISDDILRKIEEDVKKELQIKNAE
              QYNKKVEPIPVKLSIPIENYDEITHYSIVRVISKEIYNKFLTNKNVSNIEIKKDEGK
              FAVNVYEIPNSIPNSK"
gene      complement(49453..52581)
              /gene="MJECL40"
CDS       complement(49453..52581)
              /gene="MJECL40"
              /note="similar to EGAD:HI1285 percent identity: 36.1;
              identified by sequence similarity; putative"
              /codon_start=1
              /transl_table=11
              /product="type I restriction enzyme"
              /db_xref="PID:g1522673"
              /translation="MPISANKYFLLVGVFIGDKMKKEAAKLNEDYVVENAAIQRLKNL
              GYSYKHGSELTPEYNERESYRADILKNRFIKAIKNINPWLTEELALKVYKTVNIDNP
              DFNMRGKIFYEMLINGVKLEFRENGEKKTRFVKLIDFENINKNEFLVANQFEVEYYE
              NGRFRRPDLVVFINGIPIAIFEKSPKSNQTAKDNDHKTKMKDIPLQLYQYAQILVV
              SDGLETKYGSPTSDWDRFFVWEGVESDDDVEVIEVDNYGNTMYKYKGNPYTSLDILLM
              GLFKKEHLIEFLEDFIHDKKKIIATYYQFYTVKKAVDRTIKSVLYGETPEDRRIGIV
              WHAQGTGSHIMLFYAKKALKQKELNYPLLVFLTDRLELEDEQLYNVFSSVFSEAERAL
              STAELQELIKKTPGGIIFATIOKFGRKSKDEHYPFLTDRNNIIIIADEAHRSHYGTLA
              QNLRKAIPNASEFLAFTATPIDYKDRSTFLVFGDYISAYPIDRAKRHGVVPIYEARL
              VELHLTNEFIDLEFDEISERVANDPETESIEKEVFAKLEKIMLTEDYLSKVSKDIIEH
              FNKRLQDFDGKAMVVTISRKVAVELYKWITKQPNAPKIAVVMSGNKSKDPEDFHPHIR
              TKKELENLAKEFKDPESDLKMVIVVDMWLTGEDVPCLHTMYFLKPMKNHSLAQAIARV
              NRVFKDKPGGLIVDYIGIADDLSKSLSKYSSEARKDLMTDIKVVIEEMKRRYEKVTSY
              FKNINYKDWKKLSSEDLSLLTVKAYQRVAKDDNTKKEFVRNVIALKKLYLLARPHPET
              IGIKDDLEFFEMIKKMIVKYSTKKIREISQDLENDIQSLISKSISAKELVDVFEMLKK
              EKPELSVLSDEFLSEIAKIEYKDYVRDVLIKILNDDIRVRMAKNPIRFKKFSERLNEV
              IEKYRIKVITTAEMIEELVNLAKEIRKAAEEGKELGTEEELAFYDLLSYPNIPLTD
              KKRVEKIAKEIARMSGYIKARDWKKKKKNLQSKIRAKLKIILMKEGIKDYSLINKISD
              DLFEYAKNLYAI"
gene      complement(52610..53908)
              /gene="MJECL41"
CDS       complement(52610..53908)
              /gene="MJECL41"
              /note="identified by GeneMark; putative"
              /codon_start=1
              /transl_table=11
              /product="M. jannaschii predicted coding region MJECL41"
              /db_xref="PID:g1522674"
              /translation="MLFMVKFRWFETEKETDIGKIPKODWDVKKIKDIGEVAGGSTPST
              KIKEYWGGDIPWTPKDLANYEYIYISRGERNITEKAVKECSLRIFPKGTILLTSRAP
```

```
IGYVAIAKNPLTNQGFRNIIPKDGVVSEYLYLFKTMSEYLKDISGGSTFPELKG
STLKEVEIPYPSPEEQKIATVLSYFDDLIENKKKQNEILEKIALELFKNWFIDFEPF
KNEEFVYNDELDKEIPKGWKRLGGDILKVSGGESLKEARVNIVPFKFTAVNHLVK
GVCIESSDFINELALKDYKMKLYNEKSIIFQKSGESLKEARVNIVPFKFTAVNHLAVI
DSSMLNEKHYFYCLLRFLLKEIVYSVKGTTLPYLKISDIENKYIIIPPQPILQKFHS
LVQPLFEKIINNQKQIMVLKKIRDALLPKLVFGELRVEEL"
```

gene   complement(54268..55944)
                /gene="MJECL42"
CDS   complement(54268..55944)
                /gene="MJECL42"
                /note="similar to EGAD:HI287 percent identity: 39.7;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="type I restriction enzyme ECORI24/3 I M protein".
                /db_xref="PID:g1522675"
                /translation="MATLDKFLSIKENDEKTKKKESKKKSSKSNKTSESLVSHDHFEL
                TPEFENQLWKVADKLRRKMEVHQYKYVVLGLIFLRALTCRFYERRKEIEEELSNPNSE
                LYTEDPELRKMLEDEDFYLSSGVLYLPKETRWDYFVENVMSPNIGEIIDTAIEILEE
                KYPDRLKDVIPKIYAQSPLDNHDYSYLINKFSEISFGKEHRVKDVFGRIYEYFLGKFT
                EVEGKLGGKFYTPRSLJRKLIVDLDVKGGSIFDPACGSGGFFVSALEKLEREGIDINE
                LSIYGQDSDPMAYRLRTKMNLIIRGAEGDIRIDDSYHDDKFMDMTFDVVANPPFNDSE
                WDANRIKPDDPRLRIGNKKVPVPPNGNANYMWILHFIYHTAPNGKAGFVMANGALSAG
                NVEGEIRKAIIENDLVYGIVACPPKLFYNVSLPVSLWFIRREKPDYMKGKVLFINAKN
                LYKQISRQNIITEHIKKIVDKFRMFESGEDEKINELGFAKVATIDEIAKNGYYLT
                PGRYVGVKIEDDGIPPEVKMKEYSEELKKLLDEEEKLRNKVKEILDALGF"
gene   complement(56184..57338)
                /gene="MJECL43"
CDS   complement(56184..57338)
                /gene="MJECL43"
                /note="identified by GeneMark; putative"
                /codon_start=1
                /transl_table=11
                /product="M. jannaschii predicted coding region MJECL43"
                /db_xref="PID:g1522676"
                /translation="MNIDIEDFEKKPLYIQRIILRNSCIEFSKMLKKYWDKNIMIGIQ
                ISEREMSFEIEERDNKNEPIKITLPEYRSKGFKWYLAYLITLEYLRILKNGGNNDIVL
                LLDDPAVYLHPNVQKSFLEKEELSKEYQILYNTHLMSLFNEEELDRVLLVYLDKENR
                TKIKRPWSNEQKDIITPIRRALGVDKILFKENLSKILFVEGISDKFILEGLKETLK
                NLKNWYIHPLSGGDKLEDNEIVKKVRLYSCLSNFEEIKYYFLLDGDKKEKFERDKISN
                KIIFLGDENQEIEDLIDKNFYLDCVLETYKRIFTHDLEKFFKKVKEIVEKLRKSKSKII
                EELNNEFRLNNLGDFSKVDVAITIKRKLYENPELANKFEKIIDCLNGKII"
gene   complement(57338..58333)
                /gene="MJECL44"
CDS   complement(57338..58333)
                /gene="MJECL44"
                /note="identified by GeneMark; putative"
                /codon_start=1
                /transl_table=11
                /product="M. jannaschii predicted coding region MJECL44"
                /db_xref="PID:g1522677"
                /translation="MFIKWGDRLMTKTLHLTKVTIRNFRSIKVAHINNIGDIAVFVGA
                NESGKSNILKALNWFGTDKPLGGDDIPVEFLGKENKEYSKKPIIEAXFEIVNKDKFKE
                KIISNIKEILEVIKVDINEDILINFEDMLNKIQFLKFEKYADGLFKTYIYDDQLEDIT
                KEFYDCFKNVISKKTPLDLFNLIYEEVLKEEVKQNIPENQISNAVANIRGNPNFNNH
                YRKILDEIRTLNTFEEFYEICKRIENIVKSIPNTSVSISIFPGRNIQLNPYNIFTEILT
                KFNSIVTIAFINLKPKFVYLDEEMELKGAVIKNTSWSNTLKEENKTLLMPDFLE"
BASE COUNT   20177 a   9030 c   7435 g 21763 t   2 others
ORIGIN

Query Match   4.6%; Score 45.6; DB 2; Length 58407;
Best Local Similarity 48.8%; Pred. No. 3;
Matches 123; Conservative   0; Mismatches 129; Indels   0; Gaps   0;

```
Qy     90 aaaaaattaccagagggcagcaaaaatgcttgaacataatattatacttactattaag 149
          ||||||   | |||          |       | ||||| |  || || ||| | |||
Db  45041 AAAAAAATAAATACTAATATCATTAATTTAAAAATTTCATTATTATTAAATTTACTATATT 45100

Qy    150 ctagatttcctcttgatcttgctcaggtttgactggagaaaatggcaggcatggatagaaa 209
          |||||  |  | ||||| ||| |||          |   ||||| ||| || ||  || |
Db  45101 CTAGAAATTTGTTATAGAGTGCTATGGTTCTTCATTAAGCAATAAGAGAAGCTTTCTAC 45160
```

```
Qy    210 cagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagacagatgggtactt 269
          ||  || ||   |   || | ||||||  |||| | ||||  ||  | ||      |
Db  45161 CATTTCCACTAAAATAAACATAAAAAATCAAACAAAAAAACAAACAAATAAAATAAG 45220

Qy    270 gaataacaacaatggcgacaataacaacaacaataatgttacaagctcattacgagataa 329
          |  ||| |   |       ||| || || || || |  |  ||       || ||| |||
Db  45221 TACAAACAAAAAATACAACAATCAAAATAAAACAACAATAAAATAAAACTTGTCAAAATAA 45280

Qy    330 ttatggaaatga 341
          ||   |  ||||
Db  45281 GTAAATTAATAA 45292
```

Search completed: July 15, 1999, 13:56:00
Job time: 3922 sec

```
                                                                    Methanococcus jann
                                                                    BglII/HpaII fragme

OM nucleic - nucleic search, using sw model

           GenCore version 4.5
       Copyright (c) 1993 - 1998 Compugen Ltd.

Run on:    July 15, 1999, 12:52:44 ; Search time 62 Seconds
                                    (without alignments)
                                    3025.271 Million cell updates/sec

Title:          US-09-156-580-1
Perfect score:  997
Sequence:       1 cccagtgccattttcctct.........aagttaaattaaaaaaaa 997

Scoring table:  IDENTITY_NUC

Searched:       240622 seqs, 9405609 residues

Database :      N_Geneseq_34:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.

                              SUMMARIES

                %
Result          Query
NO.     Score   Match  Length DB  ID           Description
-----   -----   -----  ------ --  --------      -----------
    1   45.6    4.6     58407  1   V21210        Methanococcus jann
    2   40.8    4.1       198  1   T17202        DNA-spanner oligon
    3   40.8    4.1       198  1   T17203        DNA-spanner oligon
    4   40.6    4.1      1690  1   Q92524        Nicotiana alata ar
    5   39.2    3.9       144  1   Q12515        CSP-2 peptide from
    6   39.2    3.9       198  1   T17204        DNA-spanner oligon
    7   39.2    3.9       198  1   T17205        DNA-spanner oligon
    8   37.6    3.8       120  1   Q12516        CSP-2 peptide from
    9   37.2    3.7      2384  1   Q68909        Human c-myc far up
   10   37.2    3.7      2381  1   Q68910        Human c-myc far up
   11   37.2    3.7      2680  1   T92160        Human deleted in p
   12   36.6    3.7      8220  1   Q83529        P. falciparum Proj
   13   36.6    3.7      8220  1   T72897        Plasmodium Proj3 g
   14   36.6    3.7     19124  1   T72882        Plasmodium var-7 g
   15   36      3.6    110000  1   V21209_08     Continuation (9 of
   16   36      3.6     64976  1   V21209_16     Continuation (17 o
   17   35.8    3.6      3157  1   Q13316        Duffy receptor gen
   18   35.8    3.6       731  1   Q31693        RsaI restriction f
   19   35.8    3.6      4486  1   Q35118        Encodes E2 protein
   20   35.8    3.6      3831  1   V52424        Streptococcus pneu
   21   35.6    3.6       117  1   Q12514        CSP-2 peptide from
 c 22   35.6    3.6     40352  1   V02032        MAGE-B cluster DNA
 c 23   35.4    3.6      9636  1   Q67190        P. falciparum tran
 c 24   35.4    3.6      7400  1   T09340        Tobacco mosaic vir
   25   35.4    3.6      1942  1   T38896        Carnation ACC synt
 c 26   35.2    3.6      5852  1   Q11710        Dictyostelium plas
 c 27   35.2    3.5       154  1   T25999        Human gene signatu
   28   35      3.5      3214  1   T62360        Human origin of re
   29   35      3.5      3214  1   T73287        Human origin of re
   30   35      3.5      2160  1   T85328        Truncated murine T
   31   34.8    3.5      1016  1   N93636        Sequence of alpha-
   32   34.8    3.5      8920  1   Q52924        Carbamoyl-phosphat
   33   34.8    3.5      1856  1   T63233        Human ubiquitin co
   34   34.8    3.5      1137  1   T63234        Human ubiquitin co
   35   34.8    3.5     12987  1   T91324        Arabidopsis thalia
   36   34.8    3.5      4061  1   V44867        Clone CT797_3 codi
   37   34.4    3.5       217  1   T20147        Human gene signatu
   38   34.4    3.5       454  1   T65069        Canine genomic mic
   39   34.2    3.4      4500  1   Q29684        CCV-6 spike gene.
 c 40   34.2    3.4      4359  1   Q50617        Canine coronavirus
 c 41   34.2    3.4      4359  1   Q52447        Canine coronavirus
 c 42   34.2    3.4      2681  1   T30870        Engineered 95 kD p
 c 43   34      3.4     19124  1   T72882        Plasmodium var-7 g
```

```
                                   ALIGNMENTS

 c  44   33.8   3.4   58407  1  V21210
    45   33.6   3.4    8700  1  Q42541

RESULT   1
V21210
ID   V21210 standard; DNA; 58407 BP.
AC   V21210;
DT   10-NOV-1998 (first entry)
DE   Methanococcus jannaschii large circular extrachromosomal element.
KW   Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW   genome; autotrophic; extrachromosomal element; identification: ds.
OS   Methanococcus jannaschii.
PN   WO9807830-A2.
PD   26-FEB-1998.
PF   22-AUG-1997; U14900.
PR   22-AUG-1996; US-024428.
PA   (GENO ) INST GENOMIC RES.
PA   (UNII ) UNIV ILLINOIS FOUND.
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI   Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
DR   WPI; 98-169145/15.
PT   Complete genome sequence of methano-genic archaeon, Methanococcus
PT   jannaschii - useful in identification of M. jannaschii genome
PT   fragment
PS   Claim 13; Page 585-600; 614pp; English.
CC   The present sequence represents the large circular extrachromosomal
CC   element sequence of the Methanococcus jannaschii circular chromosome. The
CC   present invention describes M. jannaschii open reading frames from the
CC   genome sequence. The invention also describes a computer based system
CC   for identifying fragments of the M. jannaschii genome that are
CC   homologous to target nucleotide sequences, comprising: (a) data storage
CC   means comprising the nucleotide sequence of the 164976, 58407 or 16550
CC   bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
CC   least 99.9% identical to it; (b) search means for comparing a target
CC   sequence to the nucleotide sequence of the data storage means to
CC   identify a homologous sequence, and (c) retrieval means for obtaining
CC   the homologous sequence. The method, which is based on whole genome
CC   random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC   of which consists of 3 physically distinct elements, a large circular
CC   chromosome (the 164976 bp sequence given in V21209), a large circular
CC   extra-chromosomal element (the 58407 bp sequence given in V21210), and a
CC   small circular extra-chromosomal element (the 16550 bp sequence given in
CC   V21211), can be used in the identification of M. jannaschii genome
CC   fragment.
SQ   Sequence   58407 BP;     20163 A;      9034 C;      7447 G;     21761 T;

Query Match             4.6%;     Score 45.6;  DB 1;  Length 58407;
Best Local Similarity  48.8%;     Pred. No. 0.06;
Matches  123;    Conservative    0;  Mismatches  129;  Indels   0;  Gaps   0;

Qy       90 aaaaattaccagagggcagcaaaaaatgcttgaacataattattatacttactattaag 149
            ||||||| || ||| |||||||||  | || || ||| |||||| | ||||||| | |||||
Db    45041 AAAAATATACTAATAAATGCTTAAAAATTTCATTATTAAAATTTAATTTACTATTATT 45100

Qy      150 ctagattccttcttgatcttgctaggttgactggagaaaatggcaggcatggatgagaaa 209
            ||| |||  ||| |||| |||| || ||  ||  |||   |||| |  ||||| ||||| |
Db    45101 CTAGAAATTTGTTATAGAGTGCTATGGTCTTCATTAAGCAATAAGAAGAAGAGCTTTCTAC 45160

Qy      210 cagtttcaacagtagtagtacttcaaaaacaaaagcatcatggcaggcaggcactt 269
            ||||  || || ||||| | || ||||| |||| || |||| ||| |||  | |||| |||
Db    45161 CATTTCCACTAAAAATACAATAAAACATAAAAATCAAACAAAAAAAAATCAAATAAG 45220

Qy      270 gaataacaacaatggcgacaataacaacaacaataatgttacagctcattacgagataa 329
            || |||| || ||||| || || ||| || ||||| |||| || ||| |||  |||  |||
Db    45221 TACAAAAAATACAATCAAAATAAAATAAAAACAACATAAAAAATAAAAACTTGTCAAATAA 45280

Qy      330 ttatggaatga 341
            || ||||
```

Db 45281 GTAAATTAATAA 45292

```
RESULT 2
ID   T17202 standard; DNA; 198 BP.
AC   T17202;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide BAM36-198.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;
KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9605847-A1.
PD   29-FEB-1996.
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3: Page 18; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-198op, RI36-198
CC   and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC   RI36-198op, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP; 121 A; 64 C; 7 G; 6 T;

Query Match        4.1%;  Score 40.8; DB 1; Length 198;
Best Local Similarity 61.1%;  Pred. No. 0.19;
Matches  66; Conservative  0; Mismatches 42; Indels  0; Gaps  0;

Qy  198 catggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 257
          ||-||||  |||||||| ||- ||| |||||||||||||||||||||||||||||| |
Db   34 CAAGGATCCCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA  93

Qy  258 gatggagtacttgaataacaacaatggcgacaataacaacaacaataa 305
          || |- ||  |||||||||  | |||| |||||||||||||| ||
Db   94 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 141

RESULT 3
ID   T17203 standard; DNA; 198 BP.
AC   T17203;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide BAM36-198op.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;
KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9605847-A1.
PD   29-FEB-1996.
```

```
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3: Page 18; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-198op, RI36-198
CC   and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC   RI36-198op, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP; 121 A; 64 C; 7 G; 6 T;

Query Match        4.1%;  Score 40.8; DB 1; Length 198;
Best Local Similarity 61.1%;  Pred. No. 0.19;
Matches  66; Conservative  0; Mismatches 42; Indels  0; Gaps  0;

Qy  198 catggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 257
          ||-||||  |||||||| ||- ||| |||||||||||||||||||||||||||||| |
Db   34 CAAGGATCCCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA  93

Qy  258 gatggagtacttgaataacaacaatggcgacaataacaacaacaataa 305
          || |- ||  |||||||||  | |||| |||||||||||||| ||
Db   94 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 141

RESULT 4
ID   Q92524 standard; DNA; 1690 BP.
AC   Q92524;
DT   05-DEC-1995 (first entry)
DE   Nicotiana alata arabinogalactan protein cDNA NaAGP1.
KW   Arabinogalactan protein; AGP; gum; thickening agent; emulsifier;
KW   adhesive; ss.
OS   Nicotiana alata.
FH   Key             Location/Qualifiers
FT   cds             60..1445
FT                   /*tag= a
FT   misc_difference 1..31
FT                   /*tag= b
FT                   /note= "PCR SEQ non-overlapping with cDNA SEQ"
PN   WO9515377-A.
PD   08-JUN-1995.
PF   01-DEC-1994; AU0744.
PR   03-DEC-1993; US-161944.
PR   18-JUL-1994; US-276452.
PA   (ALBR ) ALBRIGHT & WILSON AUSTRALIA.
PI   Bacic A, Chen C, Clarke AE, Du H, Gane AM, Mau S;
DR   WPI; 95-215258/28.
DR   P-PSDB; R75506.
PT   Recombinant plant arabinogalactan protein and cloned DNA encoding it
PT   useful as an emulsifying agent, adhesive agent or a lubricating agent.
PS   Claim 8; Figure 1F; 142pp; English.
CC   The N-terminal region of an arabinogalactan protein (AGP) was obtd.
CC   from N. alata suspension culture. It was used as a template for
```

CC   the synthesis of an oligo primer which was further utilised for the
CC   isolation of a hybridising AGP gene from both N. alata and N.
CC   plumbaginafolia. Degenerate reverse primers corresp. to part of the
CC   AGP N-terminal AA sequence were synthesised and used to obtain a
CC   160 bp primer extension product, which was then amplified by PCR.
CC   Additional primers, corresp. in sequence to parts of the 160-bp
CC   fragment were synthesised and used to amplify the 3' part of the
CC   AGP gene by nested PCR. The primers were Q92536 and Q92537.A
CC   1.6 kb fragment was amplified and sequenced. The alignment of the
CC   sequences obtd. from the two PCR reactions gave rise to a DNA
CC   sequence shown. The PCR fragment encodes a protein contg.
CC   the peptide (R75513) by protein sequencing with two mismatches
CC   (see Q92524 FT). The 1.6 kb PCR fragment was used to screen a cDNA
CC   library made from RNA isolated from N. alata cells. The alignment
CC   of the PCR sequences with the cDNA sequence gave rise to Q92524.
CC   The calc. mol. wt. of the encoded protein is 51.8 kD with a
CC   predicted pI of 3.84.
SQ   Sequence 1690 BP;   619 A;   352 C;   274 G;   445 T;

Qy   198 catggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 257
Db   1022 CAATAACAATATGTTTTCTCCGAGAATTACAAACAACAATAACAACAATGCTTTCTA 1081

Qy   258 gatgggagtacttgaataacaacaaatggcgacaataacaacaataatgttacaagct 316
Db   1082 CGAGAATTACAACAACAACAACAACAATGGCTACTCAGAGAACTACAATCAGGCTAGCAGCT 1140

Query Match   4.1%;   Score 40.6; DB 1; Length 1690;
Best Local Similarity 58.8%;   Pred. No. 0.37;
Matches 70; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

RESULT 5
Q12515
ID   Q12515 standard; DNA; 144 BP.
AC   Q12515;
DT   18-SEP-1991 (first entry)
DE   CSP-2 peptide from P. falciparum (clone 4).
KW   P. berghei; Pf42; Pf54; Pb42; Pb54; ss.
OS   Plasmodium falciparum.
PN   WO9108756-A.
PD   27-JUN-1991.
PF   12-DEC-1990; U07194.
PR   12-DEC-1989; US-448109.
PA   (BIOM-) BIOMEDICAL RES INST.
PI   Hollingdal MR, Sina B;
DR   WPI; 91-207865/28.
DR   P-PSDB; R12903.
PT   Sporozoite polypeptide antigens - are useful for antimalarial
PT   vaccines and the encoding these peptide(s) can be incorporated
PT   into virus to produce live vaccines
PS   Disclosure: Page 21; 32pp; English.
CC   A portion of the gene encoding the P. falciparum sporozoite
CC   antigenic protein CSP-2 (mol. wt. 42kD/54kD) was cloned. Clones 3,
CC   4 and 6 were identified and found to contain malaria DNA inserts as
CC   represented in Q12514, Q12515 and Q12516 respectively. The
CC   polypeptides are present on the surface of the sporozoites and are
CC   cross-reactive with P. berghei antiserum. The Pf and Pb
CC   proteins are substantially homologous, if not identical.
SQ   Sequence 144 BP;   87 A;   48 C;   9 G;   0 U;

Qy   198 catggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 257
Db   30 CAGCAACAGCAACAACAACAACAACAACAACAACAACAACAGCAACAACAACAACAACAACACAA 89

Qy   258 gatgggagtacttgaataacaacaaatggcgacaataacaacaacaataa 305

Query Match   3.9%;   Score 39.2; DB 1; Length 144;
Best Local Similarity 60.2%;   Pred. No. 0.42;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db   90 CAACAACAACACAGCAACAACAACAACAGCAACACAACAACAACAACA 137

RESULT 6
T17204
ID   T17204 standard; DNA; 198 BP.
AC   T17204;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide RI36-198.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;
KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9605847-A1.
PD   29-FEB-1996.
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3: Page 19; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-198op, RI36-198
CC   and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC   RI36-198op, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised
SQ   Sequence 198 BP;   122 A;   63 C;   6 G;   7 T;

Qy   198 catggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 257
Db   10 CAACAACAACAACAACAACAACAACAAGAATTCCAACAACAACAACAACAACAACAACAA 69

Qy   258 gatgggagtacttgaataacaacaacaataacaacaacaataa 305
Db   70 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 117

Query Match   3.9%;   Score 39.2; DB 1; Length 198;
Best Local Similarity 60.2%;   Pred. No. 0.46;
Matches 43; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

RESULT 7
T17205
ID   T17205 standard; DNA; 198 BP.
AC   T17205;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide RI36-198op.
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;
KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9605847-A1.
PD   29-FEB-1996.

PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3; Page 19; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-1980p, RI36-198
CC   and RI36-1980p, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-1980p and
CC   RI36-1980p, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP; 122 A; 63 C; 6 G; 7 T;

Query Match
Best Local Similarity   60.2%;   Pred. No. 0.46;   Score 39.2;   DB 1;   Length 198;
Matches 65;   Conservative 0;   Mismatches 0;   Indels 43;   Gaps 0;

Qy 198 catggatagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagaca 257
Db   10 CAACACAACAACAACAACAACAACAACAAGAATTCCAACAACAACAACAACAACAACAA 69

Qy 258 gatggagtacttgaataacaacaatggcgacaataacaacaacaacaacaataa 305
Db   70 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 117

RESULT 8
Q12516
ID   Q12516 standard; DNA; 120 BP.
AC   Q12516;
DT   18-SEP-1991 (first entry)
DE   CSP-2 peptide from P. falciparum (clone 6).
KW   P. berghei Pf42; Pf54; Pb42; Pb54; ss.
OS   Plasmodium falciparum.
PN   WO9108756-A.
PD   27-JUN-1991.
PF   12-DEC-1990; U07194.
PR   12-DEC-1989; US-448109.
PA   (BIOM-) BIOMEDICAL RES INST.
PI   Hollingdal MR, Sina B;
DR   WPI; 91-207865/28.
DR   P-PSDB; R12904.
PT   Sporozoite polypeptide antigens - are useful for antimalarial
PT   vaccines and the encoding these peptide(s) can be incorporated
PT   into virus to produce live vaccines
PS   Disclosure; Page 22; 32pp; English.
CC   A portion of the gene encoding the P. falciparum sporozoite
CC   antigenic protein CSP-2 (mol. wt. 42kD/54kD) was cloned. Clones 3,
CC   4 and 6 were identified and found to contain malaria DNA inserts as
CC   represented in Q12514, Q12515 and Q12516 respectively. The
CC   polypeptides are present on the surface of the sporozoites and are
CC   cross-reactive with P. berghei antiserum. The Pf and Pb
CC   proteins are substantially homologous, if not identical.
SQ   Sequence 120 BP; 70 A; 40 C; 10 G; 0 U;

Query Match
Best Local Similarity   59.3%;   Pred. No. 0.98;   Score 37.6;   DB 1;   Length 120;
Matches 64;   Conservative 0;   Mismatches 0;   Indels 44;   Gaps 0;

Qy 208 aacagtttcaacagtagtacttcaaaaacaaaagcatcatggcaagacagatggagtac 267
Db   1 AGCAGCAACAACAACAACAGCAGCAACAACAACAGCAACAACAACAACAACAACAAC 60

Qy 268 ttgaataacaacaatggcgacaataacaacaacaacaataagttacaagc 315
Db   61 AACAACAACAACAACAGCAGCAACAACAACAACAACAACAACAACACAGC 108

RESULT 9
Q68909
ID   Q68909 standard; cDNA; 2384 BP.
AC   Q68909;
DT   13-APR-1995 (first entry)
DE   Human c-myc far upstream element (FUSE) binding protein (FBP).
KW   DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
KW   promoter P1; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   cds   27..1964
FT   /*tag= a
FT   misc_difference 473
FT   /*tag= b
FT   /label= A,G
PN   WO9419465-A.
PD   01-SEP-1994.
PF   22-FEB-1994; U01782.
PR   22-FEB-1993; US-021608.
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI   Avigan MI, Duncan RC, Levens DL;
DR   WPI; 94-294330/36.
DR   P-PSDB; R58813.
PT   New DNA-binding regulator of c-myc expression and its cDNA - used
PT   to develop prods. for diagnosis and therapy of disease states
PT   such as tumour formation
PS   Claim 1; Page 64-66; 94pp; English.
CC   DRONE (DNA-binding regulator of c-myc expression) and FUSE (far
CC   upstream element) binding protein (FBP) are synonymous. A FUSE which
CC   is required for maximal transcription of c-myc binds a factor (DROME
CC   or FBP). The activator cis-element is approx. 1500 bp 5' of the
CC   human c-myc promoter P1. A full length FBP cDNA sequence was
CC   assembled from overlapping clones obtd. from cDNA libraries. Source
CC   RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
CC   BJAB, and PMA/PHA stimulated pooled human peripheral blood
CC   lymphocytes (PBLs). A composite cDNA is given in Q68909 and the
CC   deduced polypeptide sequence. (67.5 kd mol. wt.) is given in R58813.
SQ   Sequence 2384 BP; 771 A; 488 C; 564 G; 560 T;

Query Match
Best Local Similarity   52.6%;   Pred. No. 2.7;   Score 37.2;   DB 1;   Length 2384;
Matches 81;   Conservative 0;   Mismatches 0;   Indels 73;   Gaps 0;

Qy 844 aattggtcgcacagcttagtcaaaactacttatgcacttaatatggcttcttgtgctat 903
Db 2210 AAATGTATTTGCAATTATGTGGTTGGTGTCTTTGTGATATAAAATGTACTTTTTCAATGTAT 2269

Qy 904 attatttattttacatgctgtatctagtttgcattttaagatttagtacctgtcag 963
Db 2270 ACTTTCACTTTCCAAATGCCTGTTTTGTGCTTTACAATAAATGATATGAAACCTCAAAAA 2329

Qy 964 attaaaagaaaacgaaagtaaattaaaaaaaaa 997
Db 2330 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2363

RESULT 10
Q68910

```
ID   Q68910 standard; cDNA; 2381 BP.
AC   Q68910;
DT   13-APR-1995   (first entry)
DE   Human c-myc far upstream element (FUSE) binding protein (FBP)
DE   variant lacking Ser 97.
KW   DNA-binding regulator; c-myc; DRONE; FUSE; activator; cis-element;
KW   promoter P1; ss.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   cds              27...1961
FT                    /*tag= a
FT   misc_difference  470
FT                    /*tag= b
FT                    /label= A,G
PN   WO9419465-A.
PD   01-SEP-1994.
PF   22-FEB-1994; U01782.
PR   22-FEB-1993; US-021608.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Avigan MI, Duncan RC, Levens DL;
DR   WPI: 94-294330/36.
DR   P-PSDB; R58814.
PT   New DNA-binding regulator of c-myc expression and its cDNA - used
PT   to develop prods. for diagnosis and therapy of disease states
PT   such as tumour formation
PS   Claim 8; Page 57-59; 94pp; English.
CC   DROME (DNA-binding regulator of c-myc expression) and FUSE (far
CC   upstream element) binding protein (FBP) are synonymous. A FUSE which
CC   is required for maximal transcription of c-myc binds a factor (DROME
CC   or FBP). The activator cis-element is approx. 1500 bp 5' of the
CC   human c-myc promoter P1. A full length FBP cDNA sequence was
CC   assembled from overlapping clones obtd. from cDNA libraries. Source
CC   RNAs were from undifferentiated HL60 cells, the B lymphoma cell line,
CC   BJAB, and PMA/PHA stimulated pooled human peripheral blood
CC   lymphocytes (PBLs). A composite cDNA is given in Q68909 and the
CC   deduced polypeptide sequence, (67.5 kd mol. wt.) is given in R58813.
CC   Three clones from a BJAB cDNA library contain the sequence from an
CC   activated human PBL cDNA library contain the sequence in Q68909.
CC   Three clones from the same BJAB library and two clones from the
CC   activated lymphocyte library are lacking bps 316,317 and 318
CC   (see Q68910).The mRNA lacking these three bps would encode a
CC   variant protein lacking Ser 97 (see R58814).
SQ   Sequence  2381 BP;  769 A;  488 C;  564 G;  559 T;

Query Match             3.7%;  Score 37.2;  DB 1;  Length 2381;
Best Local Similarity   52.6%;  Pred. No.2.7;
Matches   81;  Conservative    0;  Mismatches  73;  Indels    0;  Gaps    0;

Qy    844  aattggtcgcagcttagtcaaactacttatgcactttaatatgcttcctgtgctat  903
           |||||  ||  || ||| |||| ||  ||||||||| |||  ||    |||
Db   2207  AAATGTGATTTGCAATTATGTGGTTGCTTGTGATATAAATGTACTTTTTCAATGTAT  2266

Qy    904  atttattttacatggctgtatctaggtttgcatttttagagatttgatacctgtcag  963
           ||  ||||| ||  | |||||||  ||||||| |  |||   |
Db   2267  ACTTTCACTTTCCAAATGCCCTGTTTTGTGCTTTACAATAAATGATATGAAACCTCAAAA  2326

Qy    964  attaaaagaaaaacgaaagttaaattaaaaaaa  997
           |  |||| |||| |||| ||||||||
Db   2327  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2360

RESULT  11
T92160/c
ID   T92160 standard; cDNA; 2680 BP.
AC   T92160;
DT   04-MAR-1998   (first entry)
DE   Human deleted in pancreatic cancer locus 4 - DPC4 - encoding cDNA.
KW   DPC4; pancreatic cancer; deleted; locus 4; diagnosis; human;
KW   tumour suppressor gene; proliferative disease; bile duct; bladder;
KW   colorectal; cancer; Crohn's disease; colitis; ss.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   CDS              129..1787
FT                    /*tag= a
FT                    /product= DPC4
PN   WO9726271-A1.
PD   24-JUL-1997.
PF   17-JAN-1997; U00827.
PR   19-JAN-1996; US-588821.
PA   (UJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI   Hahn SA, Kern SE;
DR   WPI: 97-385290/35.
DR   P-PSDB; W14287.
PT   Deleted in Pancreatic Cancer locus 4 polypeptide - and related
PT   nucleic acids, used in diagnosis and treatment of proliferative
PT   diseases, e.g. cancer of pancreas or other organs
PS   Claim 4; Page 65-67; 104pp; English.
CC   The present sequence encodes the pure DPC4 (deleted in pancreatic
CC   cancer, locus 4) polypeptide. DPC4 is a tumour suppressor gene.
CC   Detection of truncated DPC4 protein, or of homozygous deletions or
CC   intragenic mutations in the encoding nucleic acid, is used to diagnose
CC   (in vivo or in vitro) proliferative diseases, especially pancreatic
CC   carcinoma, bile duct, bladder or colorectal cancer, Crohn's disease,
CC   colitis-associated neoplasia or chronic ulcerative colitis. These
CC   conditions, where associated with a homozygous deletion, can be treated
CC   by administering an agent that: (a) modulates DPC4 expression,
CC   specifically a sense DPC4 sequence (particularly in the form of a
CC   vector, i.e. by gene therapy), but also an antisense sequence where
CC   DPC4 protein is over expressed or (b) mimics the activity of DPC4. DPC4
CC   nucleic acid is also used as hybridisation probes for detecting
CC   presence/absence of human chromosome 18q21.1 fragments. When a
CC   homozygous deletion is detected in this region, an agent can be
CC   administered that accumulates within, or kills, only cells which
CC   contain such a deletion. This agent exploits the absence of an enzyme
CC   (or other protein) encoded by a neighbouring gene and lost by the
CC   deletion, i.e. it has a highly selective action.
SQ   Sequence  2680 BP;  775 A;  548 C;  567 G;  790 T;

Query Match             3.7%;  Score 37.2;  DB 1;  Length 2680;
Best Local Similarity   50.6%;  Pred. No. 2.8;
Matches   90;  Conservative    0;  Mismatches  88;  Indels    0;  Gaps    0;

Qy    115  aatgcttgaacataattattatacttactattaagctagattcctcttgacctgctag  174
Db   2329  AAAGATTACAATACATTATTACTTCATTTTTAAACACAGAGTAAAGTTACTCTTGGTAA  2270

Qy    175  gtttgactggagaaaatggcaggcatggatagaaacagtttcaacagtaagtacttcaaa  234
Db   2259  AATTAACTACCACCCACATGGCACCATTATTTTGAAGTATGTTCAACAATAACAACTTGTAC  2210

Qy    235  aacaaaagcatcatggcagacagcatggagtacttgaataacaacaatggcgacaata  292
Db   2209  ACAAAATAACATAGGTGAATTAACTTTGAGAACTTCTCTGCGCCTTTAAAATACGCAATA  2152

RESULT  12
Q83529
ID   Q83529 standard; DNA; 8220 BP.
AC   Q83529;
DT   22-SEP-1995   (first entry)
DE   P. falciparum Proj3 gene.
KW   Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
KW   vaccine; ss.
OS   Plasmodium falciparum.
PN   WO9507353-A.
PD   16-MAR-1995.
PF   07-SEP-1994; U10230.
PR   10-SEP-1993; US-119677.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI   Wellems TE;
DR   WPI: 95-123427/16.
DR   P-PSDB; R70236.
PT   New erythrocyte binding domain polypeptide(s) - isolated from
```

PT   Plasmodium binding proteins, used in diagnosis, treatment and
PT   prevention of malaria
PS   Disclosure: Page 61-65; 81pp; English.
CC   Erythrocyte binding ligand (EBL) family genes were cloned from
CC   P. falciparum chromosome 7 subsegment libraries constructed during
CC   genetic studies of the chloroquine resistance locus. The 4 genes,
CC   EBL-el (Q83526), E31a (Q83527), EBL-e2 (Q83528) and Proj3 (Q83529),
CC   encode the proteins given in R70233-36, respectively. The binding
CC   domains of such proteins can be expressed e.g. in E. coli, yeast,
CC   mammalian, insect, and in vaccinia virus and adenovirus-infected
CC   cells, and provide protection against P. falciparum.
SQ   Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T;

Query Match   3.7%; Score 36.6; DB 1; Length 8220;
Best Local Similarity 55.0%; Pred. No. 5.1;
Matches   72; Conservative   0; Mismatches 59; Indels   0; Gaps   0;

Qy   867 aactacttatgcacttaatatggccttcctttgtgctatatttatttttacatggctgt 926
           || | | | || | | ||| | | || | ||||||||||| |||||||| ||
Db   8030 AAGGATATATATATATGGATGTACTTGTATGTGTTTTATATATATTTATATATATGT 8089

Qy   927 atctaggtttgcatttaagattagtacctgtcagattaaaagaaacgaaagttaaa 986
           | | | | ||| | | | || | ||| | ||||| ||| | ||||| |||
Db   8090 ATTTATTATAAAAAGAAATATAAAAACAAATTATTAAAAAATGAAAAAAATGAA 8149

Qy   987 ttaaaaaaaa 997
           || ||||||||
Db   8150 ATATAAAAAA 8160

RESULT 13
T72897
ID   T72897 standard; cDNA; 8220 BP.
AC   T72897;
DT   07-OCT-1997 (first entry)
DE   Plasmodium Proj3 gene.
KW   DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW   Plasmodium; ss.
OS   Plasmodium falciparum.
FH   Key   Location/Qualifiers
FT   cds   2..8113
FT   /*tag= a
FT   /note= "encodes residues 8-2710 of W22482"
PN   WO9640766-A2.
PD   19-DEC-1996.
PF   07-JUN-1995; U09508.
PR   07-JUN-1995; US-487826.
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI   Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI   Wellems TE;
DR   WPI; 97-052231/05.
DR   P-PSDB; W22482.
PT   New malaria vaccines - contains cysteine-rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
PS   Disclosure: Page 48-50; 96pp; English.
CC   This sequence represents the Proj3 gene of Plasmodium. Proj3 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC   family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and SABP
CC   mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the invention.
CC   The compositions are for the treatment and prevention of malaria, and
CC   comprise either a nucleotide sequence or encoded polypeptide of the
CC   var-1, var-2 or var-7 genes of the DBL gene family, a family of

CC   genes having homology with conserved regions of DABP and SABP. The
CC   compositions are used for the treatment and prevention of malaria. They
CC   are also used in the preparation of vaccines for inducing a protective
CC   immune response in a mammal to Plasmodium merozoites (especially
CC   Plasmodium falciparum or Plasmodium vivax).
SQ   Sequence 8220 BP; 3194 A; 1249 C; 1777 G; 2000 T;

Query Match   3.7%; Score 36.6; DB 1; Length 8220;
Best Local Similarity 55.0%; Pred. No. 5.1;
Matches   72; Conservative   0; Mismatches 59; Indels   0; Gaps   0;

Qy   867 aactacttatgcacttaatatggccttcctttgtgctatatttatttttacatggctgt 926
           || | | | || | | ||| | | || | ||||||||||| |||||||| ||
Db   8030 AAGGATATATATATATGGATGTACTTGTATGTGTTTTATATATATTTATATATATGT 8089

Qy   927 atctaggtttgcatttaagattagtacctgtcagattaaaagaaacgaaagttaaa 986
           | | | | ||| | | | || | ||| | ||||| ||| | ||||| |||
Db   8090 ATTTATTATAAAAAGAAATATAAAAACAAATTATTAAAAAATGAAAAAAATGAA 8149

Qy   987 ttaaaaaaaa 997
           || ||||||||
Db   8150 ATATAAAAAA 8160

RESULT 14
T72882
ID   T72882 standard; cDNA; 19124 BP.
AC   T72882;
DT   12-SEP-1997 (first entry)
DE   Plasmodium var-7 gene.
KW   DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW   Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW   DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW   Plasmodium; ss.
OS   Plasmodium vivax.
OS   Plasmodium falciparum.
FH   Key   Location/Qualifiers
FT   exon   7317..15139
FT   /*tag= a
FT   /number= 1
FT   intron   15140..16205
FT   /*tag= b
FT   /number= 1
FT   exon   16206..17552
FT   /*tag= c
FT   /number= 2
FT   /note= "no stop codon given"
PN   WO9640766-A2.
PD   19-DEC-1996.
PF   07-JUN-1996; U09508.
PR   07-JUN-1995; US-487826.
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI   Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI   Wellems TE;
DR   WPI; 97-052231/05.
DR   P-PSDB; W22482.
PT   New malaria vaccines - contains cysteine-rich DBL family protein
PT   binding domains homologous domains of the Duffy and sialic acid
PT   binding proteins
PS   Claim 4; Page 56-61; 96pp; English.
CC   This sequence represents the var-7 gene of plasmodium. var-7 belongs to
CC   the Duffy binding like (DBL) family of genes which have homology to the
CC   Duffy antigen binding protein (DABP) and sialic acid binding protein
CC   (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC   family of genes modulate cytoadherence and antigenic variation of
CC   Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC   protein (DABP) are soluble proteins that appear in the culture
CC   supernatant after infected erythrocytes release merozoites. DABP and SABP
CC   mediate the binding of merozoites and schizonts to the erythrocyte
CC   surface. These proteins are necessary for erythrocyte invasion by the
CC   parasite. This sequence can be used in the compositions of the invention.
CC   The compositions are for the treatment and prevention of malaria, and

CC   comprise either a nucleotide sequence or encoded polypeptide of the
CC   var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC   genes having homology with conserved regions of DABP and SABP. The
CC   compositions are used for the treatment and prevention of malaria. They
CC   are also used in the preparation of vaccines for inducing a protective
CC   immune response in a mammal to Plasmodium merozoites (especially
CC   Plasmodium falciparum or Plasmodium vivax).
SQ   Sequence   19124 BP;   2190 A;   2790 C;   2790 G;   6320 T;

Query Match   3.7%;   Score 36.6;   DB 1;   Length 19124;
Best Local Similarity   55.0%;   Pred. No. 6.4;
Matches   72;   Conservative   0;   Mismatches   59;   Indels   0;   Gaps   0;

Qy   867 aactacttatgcacttaatatggcttccttgtgctatatttattttacatggctgt 926
          || | ||| || | | |||| | ||| || || ||||||| |||| ||| | ||
Db 15350 AAGGATATATATATATGGATGTACTTGTATGTGTTTTATATATATATTTATATATGT 15409

Qy   927 atctaggtttgcatttaagattagtacctgtacctgtcagattaaaagaaacgaaagttaaa 986
          || ||| || | ||| | || || || |||||| | ||||||| ||| | || | ||
Db 15410 ATTTATTAAAAAAGAAATATAAAAACAAATTATTAAAATCAAAAAAAGAAAAAATGAA 15469

Qy   987 ttaaaaaaaa 997
          ||| |||||||
Db 15470 ATATAAAAAA 15480

RESULT 15
V21209_08/c
Continuation (9 of 17) of V21209 from base 800001 (Methanococcus jannaschii circular chr
WP Sequence split into 17 fragments   LOCUS V21209   Accession V21209
WP   Fragment Name   Begin   End
WP   V21209_00   1   110000
WP   V21209_01   100001   210000
WP   V21209_02   200001   310000
WP   V21209_03   300001   410000
WP   V21209_04   400001   510000
WP   V21209_05   500001   610000
WP   V21209_06   600001   710000
WP   V21209_07   700001   810000
WP   V21209_08   800001   910000
WP   V21209_09   900001   1010000
WP   V21209_10   1000001   1110000
WP   V21209_11   1100001   1210000
WP   V21209_12   1200001   1310000
WP   V21209_13   1300001   1410000
WP   V21209_14   1400001   1510000
WP   V21209_15   1500001   1610000
WP   V21209_16   1600001   1664976

Query Match   3.6%;   Score 36;   DB 1;   Length 110000;
Best Local Similarity   56.9%;   Pred. No. 14;
Matches   66;   Conservative   0;   Mismatches   50;   Indels   0;   Gaps   0;

Qy   882 ttaatatggctccttgtgctatatttattttattttacatggctgtatccaggtttgcatt 941
          | ||| || || | ||||| || | | ||||||| |||| || |||| || ||| |||
Db 87972 TCAATTCTTATTTTCTTTCTTTTTTATTTTATTTTGAAACTATATTTGTAAATTTAAATT 87913

Qy   942 ttaagattagtacctgtcagattaaaagaaacgaaagttaaattaaaaaaaa 997
          | | ||| || || | | ||||| || ||| || | ||| ||||||||
Db 87912 TACAAATATAGAAACTTTAAATAACAATTTAAAACAAAATAAGAATTGAACAAATA 87857

Search completed: July 15, 1999, 13:58:17
Job time: 3933 sec

```
        GenCore version 4.5
        Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 11:12:38 ; Search time 377.02 Seconds
                                   (without alignments)
                                   4033.866 Million cell updates/sec

Title:         US-09-156-580-1
Perfect score: 997
Sequence:      1 cccagtgccattttctct.........aaagttaaattaaaaaaaa 997
Scoring table: IDENTITY_NUC
Searched:      2002476 seqs, 762712212 residues
Database :     EST:*
               1: em_est1:*
               2: em_est2:*
               3: em_est3:*
               4: em_est4:*
               5: em_est5:*
               6: em_est6:*
               7: em_est7:*
               8: em_est8:*
               9: em_est9:*
              10: gb_est1:*
              11: gb_est2:*
              12: gb_est3:*
              13: gb_est4:*
              14: gb_est5:*
              15: gb_est6:*
              16: gb_est7:*
              17: gb_est8:*
              18: gb_est9:*
              19: gb_est10:*
              20: gb_est11:*
              21: gb_est12:*
              22: gb_est13:*
              23: gb_est22:*
              24: gb_est14:*
              25: gb_est15:*
              26: gb_est16:*
              27: gb_est17:*
              28: gb_est18:*
              29: gb_est19:*
              30: gb_est20:*
              31: gb_est21:*
              32: em_est10:*
              33: em_est11:*
              34: em_est12:*
              35: em_est13:*
              36: em_est14:*
              37: em_est15:*
              38: em_est16:*
              39: em_est17:*
              40: em_est18:*
              41: em_est19:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB ID		Description
c	1	45.4	243	28	C92842	Dictyosteli	C92842 Dictyosteli
	2	42.4	303	31	AU034213	Dictyoste	AU034213 Dictyoste

No.		Score				ID	Description
3	c	42.2	4.2	482	28	C91001	C91001 Dictyosteli
4		42	4.2	605	31	AI227944	AI227944 EST224639
5	c	41.8	4.2	419	31	AU033391	AU033391 Dictyoste
6		41.4	4.2	358	27	AA634656	AA634656 ab26b00.r
7		41.4	4.2	591	31	AU033325	AU033325 Dictyoste
8	c	40.8	4.1	404	31	AA225651	AA225651 nc08b08.r
9	c	40.8	4.1	291	26	AA810593	AA810593 ob64b04.s
10	c	40.6	4.1	554	19	C23753	C23753 Dictyosteli
11		40.6	4.1	274	28	C91401	C91401 Dictyosteli
12	c	40.4	4.1	294	31	AU034848	AU034848 Dictyoste
13	c	39.8	4.0	560	23	AI199031	AI199031 q139h04.x
14		39.8	4.0	574	28	C91220	C91220 Dictyosteli
15		39.4	4.0	546	27	C22974	C22974 Dictyosteli
16		39.4	4.0	312	23	AI163084	AI163084 A031p65u
17		39.4	4.0	397	31	AU034056	AU034056 Dictyoste
18		39.2	3.9	472	12	N39467	N39467 yw62d10.r1
19		39.2	3.9	389	15	SSC5A01	Z81149 S.scrofa mR
20		39.2	3.9	226	28	C90268	C90268 Dictyosteli
21		39.2	3.9	552	28	C91403	C91403 Dictyosteli
22		39.2	3.9	480	28	C93684	C93684 Dictyosteli
23		39.2	3.9	605	31	AU033779	AU033779 Dictyoste
24	c	39	3.9	402	31	R99317	R99317 yq71h03.s1
25		39	3.9	249	28	C94333	C94333 Dictyosteli
26		39	3.9	246	31	AU033454	AU033454 Dictyoste
27		39	3.9	617	31	AU033655	AU033655 Dictyoste
28		38.8	3.9	198	31	AU034673	AU034673 Dictyoste
29		38.6	3.9	466	20	C25660	C25660 Dictyosteli
30		38.4	3.9	400	27	C84210	C84210 Dictyosteli
31		38.4	3.8	408	31	AU033476	AU033476 Dictyoste
32	c	38.2	3.8	355	15	N98120	N98120 2212C3 czap
33	c	38.2	3.8	250	21	AA278587	AA278587 zs77g03.r
34		38.2	3.8	440	28	C89946	C89946 Dictyosteli
35		38.2	3.8	497	28	C94312	C94312 Dictyosteli
36	c	38.2	3.8	529	30	AI089028	AI089028 oy61g05.s
37	c	38.2	3.8	592	31	AI146745	AI146745 qb92d08.x
38	c	38.2	3.8	526	31	AU033912	AU033912 Dictyoste
39	c	38	3.8	395	25	AA682058	AA682058 vu13c07.s
40		38	3.8	660	28	AA800394	AA800394 EST189891
41		38	3.8	643	28	C90989	C90989 Dictyosteli
42		38	3.8	627	28	C91439	C91439 Dictyosteli
43		37.8	3.8	533	11	H17378	H17378 ym38d12.r1
44	c	37.8	3.8	455	24	AA602349	AA602349 no89h11.s
45		37.8	3.8	592	28	C89805	C89805 Dictyosteli

ALIGNMENTS

```
RESULT 1
C92842/c
LOCUS       C92842    243 bp    mRNA        EST        03-JUN-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSF465, mRNA sequence.
ACCESSION   C92842
NID         g3074718
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 243)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 243)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES    Location/Qualifiers
```

```
source          1..243
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="SS"
                /dev_stage="slug"
BASE COUNT      117 a        8 c       17 g      101 t
ORIGIN

Query Match     4.6%;     Score 45.4; DB 28; Length 243;
Best Local Similarity 52.4%;  Pred. No. 0.38;
Matches 100; Conservative  0; Mismatches 91; Indels  0; Gaps   0;

Qy  807 ttaactatatcaattttacggcattaaggttgtaaattgagtcgacagcttagtcaa 866
        |||   || |||||| |||||||| |||||||||||||||||||||| |
Db  243 TTTTTTTTTTTTTTCTCTATTTTTTTTTTTTTTTTTTTTTTTTTCATAATTTCTTCAA 184

Qy  867 aactactatgcacttaaatatggcttcttggctatattattttacatggctgt 926
        ||||||| ||   |||||| ||||||||||||||||||  ||
Db  183 TTCCCAAAACTGTTTTAAAATATATTTTTGTATTAAAAATTTTAAAAAATATTACAGT 124

Qy  927 atctaggttgcatttaagattagtaccctgtcaggttaaaagaaacgaaagttaa 986
        |||||||||  | |||||  |||  |||| ||||||| ||| |||
Db  123 ACCTATCTATTCATTTATTATTTTTGTAAAATTAAAGAAAAAAAAAGAAAGAAAA 64

Qy  987 ttaaaaaaaa 997
        | |||| |||
Db   63 TAAAAATAAA 53


RESULT  2
LOCUS       AU034213    303 bp   mRNA                         EST   28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLC264, mRNA sequence.
ACCESSION   AU034213
NID         g3799637
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostellida; Dictyostelium.
REFERENCE   1  (bases 1 to 303)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 303)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..303
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      172 a       67 c       12 g       52 t
ORIGIN

Query Match     4.3%;     Score 42.4; DB 31; Length 303;
Best Local Similarity 58.9%;  Pred. No. 1.9;
Matches 73; Conservative  0; Mismatches 51; Indels  0; Gaps   0;

Qy  198 catggtagaaacagttcaacagtaagtacttcaaaaacaaagcatcatggcaagaca 257
        ||  |  || |||||||| ||||||||||||||| |||||||  |||||||| ||
Db   54 CAACAACAACCACTACAACAACCACTACAACAACAACAACAACAACAACAACAACAA 113
```

```
Qy  258 gatggagtacttgataacaacaacatggcgacaataacaacaacaataatgttacaagctc 317
        | || || |||||||||||||||||||||| | |||| |||||||||||||| |||
Db  114 CAACAACAACAACAACAACAACAACAACAACCACCAACAACAACAATATTATTACCAA 173

Qy  318 atta 321
        | ||
Db  174 ACTA 177


RESULT  3
LOCUS       C91001      482 bp   mRNA                         EST   15-APR-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSJ551, mRNA sequence.
ACCESSION   C91001
NID         g3060367
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostellida; Dictyostelium.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 482)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..482
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="slug"
BASE COUNT      232 a       98 c       51 g      101 t
ORIGIN

Query Match     4.2%;     Score 42.2; DB 28; Length 482;
Best Local Similarity 61.3%;  Pred. No. 2.2;
Matches 68; Conservative  0; Mismatches 43; Indels  0; Gaps   0;

Qy  198 catggtagaaacagttcaacagtaagtacttcaaaaacaaagcatcatggcaagaca 257
        ||  |||| ||||||| |||||||||||||| |||||||| | |||||||||
Db  368 CAGCAACAGCAACAATAACAACAACAACAACAACAACAACAACAACAACAACAACAA 427

Qy  258 gatggagtacttgataacaacaacatggcgacaataacaacaacaataatgttacaagctc 308
        | || || ||||||||||||||||||||||||| |||||||||||||
Db  428 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAT 478


RESULT  4
AI227944/c
LOCUS       AI227944    605 bp   mRNA                         EST   30-OCT-1998
DEFINITION  EST224639 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRCN90 3' end similar to similar to guanine nucleotide
            dissociation stimulator a ras-related GTPase, mRNA sequence.
ACCESSION   AI227944
NID         g3811831
KEYWORDS    EST.
SOURCE      Rattus sp.
ORGANISM    Rattus sp.
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush.J.,
```

TITLE   Kerlavage,A.R. and Adams,M.D.
           Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
           Gene Index
JOURNAL   Unpublished (1998)
COMMENT   Other_ESTs: TC57830
           Contact: Lee, NH
           ATCC
           The Institute for Genomic Research
           9712, Medical Center Drive, Rockville, MD 20850, USA
           Tel: (301)-838-3529
           Fax: (301)-838-0208
           Email: nhlee@tigr.org
           Seq primer: M13-21.
FEATURES   Location/Qualifiers
   source   1..505
                 /organism="Rattus sp."
                 /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
                 Site_2: NotI"
                 /db_xref="taxon:10118"
                 /clone="RBRCN90"
                 /clone_lib="Normalized rat brain, Bento Soares"
BASE COUNT   129 a   143 c   166 g   167 t
ORIGIN

Query Match   4.2%; Score 42; DB 31; Length 605;
Best Local Similarity 57.7%; Pred. No. 2.5;
Matches 75; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy   866 aaactacttatgcactttaatatggcttcttgtgctatatttattttacatggctg 925
             ||| |||||| || || ||||||| | |||||||||||| || |||
Db   141 ATAAAACGTAAAAGTTTACTGATTGGGGTGGGGCAATATATTTATTTCTTGTAAATAGAAA 82

Qy   926 tatctggtttgcatttagattagtacctgtcagattaaagagaaacgaaagttaa 985
             |||| ||||| ||| | |||||||| | |||||||||||| ||||| ||||
Db   81 ATGCTAGACTTGAATATATATTAAAATCCCGTTTCTACTAAAAAAAAAAAAAAAAA 22

Qy   986 attaaaaaa 995
             | ||||||
Db   21 AAAAAAAA 12

RESULT   5
AU033391
LOCUS   AU033391   419 bp   mRNA   EST   28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLA713, mRNA sequence.
ACCESSION   AU033391
NID   g3798815
KEYWORDS   EST.
SOURCE   Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM   Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 419)
AUTHORS   Urushihara,H.
TITLE   Developmental cDNA in Dictyostelium discoideum
JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 419)
AUTHORS   Urushihara,H.
TITLE   Direct Submission
JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT   PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES   Location/Qualifiers
   source   1..419
                 /organism="Dictyostelium discoideum"
                 /strain="AX4"
                 /db_xref="taxon:44689"
                 /clone_lib="SL"
                 /dev_stage="slug"

BASE COUNT   248 a   91 c   24 g   56 t
ORIGIN

Query Match   4.2%; Score 41.8; DB 31; Length 419;
Best Local Similarity 61.5%; Pred. No. 2.7;
Matches 67; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy   198 catggatagaaacagttcaacagtagtacttcaaaacaaagcatcatggcaagaca 257
             | ||||| |||||| | ||||| |||||||||| ||| |||| |
Db   191 CAGCAACAGCCAATAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 250

Qy   258 gatgggtacttgaataacaacaatggcgacaataacaacaacaacaataat 306
             || | |||||||||||| ||||| ||||||||||||||||||||
Db   251 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAT 299

RESULT   6
AA634656/c
LOCUS   AA634656   358 bp   mRNA   EST   06-MAR-1998
DEFINITION ab26b06.r1 Stratagene lung (#937210) Homo sapiens cDNA clone 841907
            5' similar to TR:G510703 G510703 LAMININ S B3 CHAIN PRECURSOR. [1]
            ;, mRNA sequence.
ACCESSION   AA634656
NID   g2557870
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 358)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE   WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Uni-ZAP XR Vector: -5; adaptor sequence: 5' GAATTCGGCACGAG
            3' -3' adaptor sequence: 5' CTCGGAGTTTTTTTTTTTTTTTTTT 3'."
            Insert Length: 759   Std Error: 0.00
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 348.
FEATURES   Location/Qualifiers
   source   1..358
                 /organism="Homo sapiens"
                 /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                 Oligo dT. normal lung.   Average insert size: 1.0 kb;
                 XR Vector: -5; adaptor sequence: 5' GAATTCGGCACGAG
                 /db_xref="taxon:9606"
                 /clone="841907"
                 /clone_lib="Stratagene lung (#937210)"
                 /sex="male"
                 /dev_stage="72 years"
                 /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT   74 a   76 c   106 g   102 t
ORIGIN

Query Match   4.2%; Score 41.4; DB 27; Length 358;
Best Local Similarity 56.1%; Pred. No. 3.4;
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy   202 gatagaaacagttcaacagtagtacttcaaaacaaagcatcatggcaagacagatg 261
             ||| ||| ||| ||| ||| ||| |||||||||||||| | ||| ||||| ||

```
Db    317  GATGAAAGAATTCAAACCCAATCTACCCCAACAACAAAAGCAAAAGCGGCAGATGAGTG 258

Qy    262  ggtacttgaataacaacaatggcgacaacaacaacaataatgttacaagctcatta 321
                |||||--|--|-|||-|| |||||    |-||||| ||||||||||||||||||||||
Db    257  GGCCAACGGGATGAAAGCTGTAGCAACACTTACAGTGCATAGTAGAGCACGCGCCACAT 198

Qy    322  cgagtaattatgggaaatg 340
                || |||| |||| |||||
Db    197  TGATATGATCACGGATCTG 179


RESULT  7
AU033325
LOCUS        AU033325       591 bp    mRNA              EST        28-OCT-1998
DEFINITION   Dictyostelium discoideum slug cDNA, clone SLA616, mRNA sequence.
ACCESSION    AU033325
NID          g3798749
KEYWORDS     EST.
SOURCE       Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
             clone_lib:SL.
ORGANISM     Dictyostelium discoideum
             Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE    1  (bases 1 to 591)
AUTHORS      Urushihara,H.
TITLE        Developmental cDNA in Dictyostelium discoideum
JOURNAL      Published Only in DataBase (1998) In press
REFERENCE    2  (bases 1 to 591)
AUTHORS      Urushihara,H.
TITLE        Direct Submission
JOURNAL      Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
             Urushihara, University of Tsukuba, Institute of Biological
             sciences; 1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
             (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
             Fax:+81-0298-53-6614)
COMMENT      PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT       234 a    115 c     77 g    162 t      3 others
ORIGIN

Query Match             4.2%;  Score 41.4; DB 31; Length 591;
Best Local Similarity   55.1%;  Pred. No. 3.5;
Matches   81;  Conservative   0; Mismatches   66; Indels   0; Gaps   0;

Qy    198  catggtgatagaaacagttcaacagtaagtacttcaaaaacaaagcatcatggccaagaca 257
                |||| |||||||||||||||||||||||||||| |||||| |||||||||||||||||||||
Db    375  CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 434

Qy    258  gatggagtacttgaataacaacaatggcgacaataacaacaacaatgttacaagctc 317
                ||||||||| ||||| |-|| |||| ||||| |||||||||||||||| ||||||||
Db    435  CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAAAACTACAACC 494

Qy    318  attacgagataattatgggaaatgaaga 344
                |-|| |  |-|||  |-|||  ||| |||
Db    495  ACAACCCTCAACCTCCATGTACATGTAGA 521


RESULT  8
AA225651/c
LOCUS        AA225651       404 bp    mRNA              EST        20-AUG-1997
DEFINITION   nc08b08.r1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:1007511, mRNA
             sequence.
ACCESSION    AA225651
NID          g1846968
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
```

```
             Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
             Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
             Homo.
REFERENCE    1  (bases 1 to 404)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
             M.D., Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html

             Insert Length: 556   Std Error: 0.00
             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 382.
FEATURES             Location/Qualifiers
     source          1..404
                     /organism="Homo sapiens"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected, histologically normal
                     prostate epithelial cells. Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies). Average insert size is 600
                     bp. NOTE: Not directionally cloned. This library was
                     constructed by David Krizman."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1007511"
                     /clone_lib="NCI_CGAP_Pr1"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     <1..>404
BASE COUNT       131 a     67 c     53 g    148 t      5 others
ORIGIN

Query Match             4.1%;  Score 40.8; DB 21; Length 404;
Best Local Similarity   49.5%;  Pred. No. 4.7;
Matches  105;  Conservative   0; Mismatches  107; Indels   0; Gaps   0;

Qy    786  tgaacttcgactgggatccacttaactatatctaattttacggcattaaggtttgtaa 845
                | || ||| ||  |||| || |-||  |-| ||| | |||||-||| |-|| ||||||
Db    249  TGCATTTGTTTGGTATTGTATTTATTCATAAAGTATTTAATTAGTGCTAAGTGTGAAC 190

Qy    846  ttggtcgacagcttagtcaaaactacttatgtcacttaatatggcttcttcttgtctatat 905
                ||| | |||| ||  |-| ||  ||| |||| ||| |-|| |||| ||| |-||||| |||
Db    189  TGGACCCTGTTGCTAAGCCCCAGCAAGCAATCCTAGGTAGGGTTTAATCCCCAGTAAAAT 130

Qy    906  ttatttatttacatggcgtatccaggtttgcatttagatttagtacttgtcagat 965
                | |-| |||-||| |-||| |-||||| || ||| |  |-|| ||  |-|||| |||
Db    129  TGCCATATTGCACATGTCTTAATGAGTTTGAATGTAAAATTGTAAATTGTATATTCACTTAA 70

Qy    966  taaagaaacgaaagtaaattaaaaaaaaa 997
                ||||||||| |||| ||||
Db    69   AAAAAAAAAAAAAAAAAACAAAAAAAAAAA 38


RESULT  9
AA810593/c
LOCUS        AA810593       291 bp    mRNA              EST        19-FEB-1998
```

DEFINITION   ob64b04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336111,
             mRNA sequence.
ACCESSION   AA810593
NID   g2880204
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 291)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
             Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
             Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/bbrp/image/image.html
             Insert Length: 926   Std Error: 0.00
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 242.
FEATURES   Location/Qualifiers
     source   1..291
                     /organism="Homo sapiens"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAGTGGGACGGCCGCCTCATTTTTTTTTTTTTTTTT-
                     3']. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1336111"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
BASE COUNT   97 a   45 c   36 g   113 t
ORIGIN

Query Match   4.1%;   Score 40.8; DB 26;   Length 291;
Best Local Similarity 49.5%;   Pred. No. 4.5;
Matches 105; Conservative   0; Mismatches 107; Indels   0; Gaps   0;

Qy   786 tgaacttcgactggatccaccttaactatctaattttacggcattaaggtttgtaa 845
             |||| ||||| ||||| |||| || || |||||||||||||||||| | |||||
Db   218 TGCATTTGTTTGGTATTGTATTTATTCAATAAAGTATTTAATTAGTGCTAAGTGTGAAC 159

Qy   846 ttgagtcgcagacgcttgtcaaaactacttatgcacttaatatggcttcttgtgctatat 905
             ||| ||||||| ||||||| ||||| || | | |||| || | |||||| |||||||
Db   158 TGGACCCTGTTGCTAAGCCCCAGCAAGCAATCCTAGGTAGGGTTAATCCCCAGTAAAAT 99

Qy   906 ttatttatttacatggctgtatctaggttgcatttaagatttagtaccttgtcagat 965
             || ||| |||| ||||| | || |||| ||| | || ||| ||| || ||| ||
Db   98 TGCCATTGCCACATGTCTTAATGAAGTTGAATGTTAAATAAATTGTATATTCACTTTA 39

Qy   966 taaagaaaacgaaagtaaagtaaattaaataaaaaaaa 997

Db   38 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 7
             |||| |||| ||||| ||| | || |||||||||||

RESULT 10
C23753/c
LOCUS   C23753   554 bp   mRNA   EST   22-APR-1997
DEFINITION Dictyostelium discoideum gamete cDNA, clone FC-AF03, mRNA sequence.
ACCESSION   C23753
NID   g2073265
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE   Dictyostelium discoideum (strain:KAX3) Sexually competent Gamete
            cDNA to mRNA, clone_lib:FC.
ORGANISM   Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyostelida;
            Dictyostelium.
REFERENCE   1 (bases 1 to 554)
AUTHORS   Urushihara,H.
TITLE   Direct Submission
JOURNAL   Submitted (04-APR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Tennohdai, Tsukuba, Ibaraki 305, Japan
            (E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
            Fax:0298-53-6614)
REFERENCE   2 (sites)
AUTHORS   Kobayashi,A., Shimizu,H. and Urushihara,H.
TITLE   Sexual cDNA in CSM
JOURNAL   Unpublished (1997)
COMMENT   PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES   Location/Qualifiers
     source   1..554
                     /organism="Dictyostelium discoideum"
                     /strain="KAX3"
                     /db_xref="taxon:44689"
                     /cell_type="Gamete"
                     /clone_lib="FC"
                     /dev_stage="Sexually competent"
BASE COUNT   117 a   36 c   123 g   278 t
ORIGIN

Query Match   4.1%;   Score 40.6; DB 19;   Length 554;
Best Local Similarity 62.1%;   Pred. No. 5.4;
Matches 64; Conservative   0; Mismatches   39; Indels   0; Gaps   0;

Qy   203 atagaaacagttcaacagtaagtacttcaaaaacaaagcatcatggcaagcaagacagatgg 262
             | |||||||||||||||| | ||| ||||||| ||||||||||||||||||||||||||||
Db   172 ACAACAACAGTATCACCAACAACAAGAGCAACAACATATTATCAACAACAACAACAACA 113

Qy   263 agtacttgaataacaacaatggcgacaataacaacaacaataa 305
             | | || ||||||||||||||||||||||||||||||||||||
Db   112 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 70

RESULT 11
C91401/c
LOCUS   C91401   274 bp   mRNA   EST   15-APR-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SSK170, mRNA sequence.
ACCESSION   C91401
NID   g3060767
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE   Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
ORGANISM   Dictyostelium discoideum
            Eukaryota; Dictyostellida; Dictyostelium.
REFERENCE   1 (bases 1 to 274)
AUTHORS   Urushihara,H.
TITLE   Developmental cDNA in Dictyostelium discoideum
JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 274)
AUTHORS   Urushihara,H.
TITLE   Direct Submission
JOURNAL   Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko

```
                Urushihara, University of Tsukuba, Institute of Biological
                Sciences: 1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
                (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
                Fax:+81-0298-53-6614)
COMMENT         PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES        Location/Qualifiers
     source     1..274
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="SS"
                /dev_stage="slug"
BASE COUNT      133 a      23 c      10 g     108 t
ORIGIN


Query Match            4.1%;   Score 40.6;   DB 28;   Length 274;
Best Local Similarity  53.5%;   Pred. No. 5;
      Matches   85;   Conservative   0;   Mismatches  74;   Indels   0;   Gaps   0;

Qy    838  tttgtaaatggagtcgacagcttagtcaaactacttatgcacttaatatggccttcttg  897
           || |||  |  |  || |||| || | | |||||||| ||  |||||
Db    272  TTTTTTTTTTTTTTTTTGATTTGTTAAAAGAAAATTTCTTTATTTTAAAATAACTAGTTG  213

Qy    898  tgctatatttattttacatggctgtcatctaggtttgcattttagtagttagtacct  957
           || |||  |  ||  ||||| | |||| | | ||||||  |||||||| | |
Db    212  TAATAGTATTTAAATTTTTACAATTGTTTTTTTTTTTTTCCAAATAAGATTTAAAATAT  153

Qy    958  tgtcagattaaaagaaaacgaaagttaaattaaaaaaa  996
           || |||||||| | ||| |||  |  ||||
Db    152  TAAAATATTAAAAGATTAAAAGATTAAAAGATTAAAGA  114


RESULT 12
AU034848
LOCUS       AU034848      294 bp    mRNA            EST        28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLE578, mRNA sequence.
ACCESSION   AU034848
NID         g3800272
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 294)
  AUTHORS   Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 294)
  AUTHORS   Urushihara,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences: 1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES    Location/Qualifiers
     source     1..294
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="SL"
                /dev_stage="slug"
BASE COUNT      176 a      52 c      14 g      52 t
ORIGIN


Query Match            4.1%;   Score 40.4;   DB 31;   Length 294;
Best Local Similarity  48.7%;   Pred. No. 5.6;
      Matches  110;   Conservative   0;   Mismatches  116;   Indels   0;   Gaps   0;

Qy     80  aaccctcacaaaaaattaccagaggggcagcaaaaatgcttgaacatataattattatact  139
```

```
Db     62  AAACAAAAGAAAAATAAAATATAAAAAACACCAAGAGAGTGATTTAAAACATATAATCAT  121
           ||| |  ||||||
Qy    140  tactattaagctagattccttcttgatccttgtccttggtttgactgggaaaatggcaggca  199
           ||  ||| || |||   |  ||| |||| | | | ||
Db    122  TAAGTAACAACACTTTATTGAACTCAGAAAACCAAACTACAACTAATTAATAACCA  181

Qy    200  tggatagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagacaga  259
           || |||  | ||| |  |||| |  || |||  |||| |||| ||||||
Db    182  ATATTAGTAATATTAATAATAAGAAGAAAAACAACAACAACAACAACAACAACAACA  241

Qy    260  tggagtacttgaataacaacaatggcgacaataacaacaacaataa  305
           || ||| |||| || ||  ||| |  ||||||||||||||
Db    242  ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA  287


RESULT 13
AI199031/c
LOCUS       AI199031       560 bp    mRNA            EST        01-DEC-1998
DEFINITION  q139h04.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1858903 3'
            similar to gb:X70218 PROTEIN PHOSPHATASE PP-X (HUMAN);contains
            element MSRI repetitive element ;, mRNA sequence.
ACCESSION   AI199031
NID         g3751637
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image.html

            Insert Length: 980    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 331.
FEATURES    Location/Qualifiers
     source     1..560
                /organism="Homo sapiens"
                /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                modified polylinker; Site_1: Not I - oligo(dT) primer (5'
                strand cDNA was primed with a Not I - oligo(dT) primer (5'
                TGTTACCAATCTGAAGTGGGAGCGGCCGCCATAGGTTTTTTTTTTTTTTTTTTT
                T 3'); double-stranded cDNA was ligated to Eco RI
                adaptors (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of the modified pT7T3 vector.
                Library is normalized, and was constructed by Bento
                Soares and M.Fatima Bonaldo."
                /db_xref="taxon:9606"
                /clone="IMAGE:1858903"
                /clone_lib="NCI_CGAP_Brn25"
                /tissue_type="anaplastic oligodendroglioma"
                /lab_host="DH10B"
BASE COUNT      151 a     129 c     150 g     130 t
ORIGIN
```

```
ORGANISM    to mRNA, clone_lib:FC.
            Dictyostelium discoideum
            Eukaryotae; mitochondrial eukaryotes; Dictyostelida;
            Dictyostelium.
REFERENCE   1 (bases 1 to 546)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; Ten-noudai 1-1-1, Tsukuba-shi, Ibaraki-ken 305, Japan
            (E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
            Fax:0298-53-6614)
REFERENCE   2 (sites)
AUTHORS     Shimizu,H. and Urushihara,H.
TITLE       Sexual-cDNA in CSM
JOURNAL     Unpublished (1997)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..546
                     /organism="Dictyostelium discoideum"
                     /strain="KAX3"
                     /db_xref="taxon:44689"
                     /cell_type="Gamete"
                     /clone_lib="FC"
                     /dev_stage="Sexually mature"
BASE COUNT      317 a    110 c     33 g     85 t      1 others
ORIGIN

Query Match              4.0%;  Score 39.4;  DB 17;  Length 546;
Best Local Similarity   57.9%;  Pred. No. 10;
Matches   70;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy   185 agaaaatggcaggcatggatgaaacagttcaacgtagtacttcaaaacaaagca 244
             ||||| ||    ||||||  |   |||||  ||||||||||||||||||||||||
Db   160 AAAAATTGAACAACAACACAACAACAACAACAACAACAACAACAACAACAACAACA 219

Qy   245 tcatggcaagacagatgggagtacttgaataacaacaatggcgaataacaacaacaata 304
           || |||||||| | |  || | |||  || |||||||||||||||||||||||||
Db   220 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 279

Qy   305 a 305
           |
Db   280 A 280

Search completed: July 15, 1999, 12:07:21
Job time: 3283 sec
```

```
Query Match              4.0%;  Score 39.8;  DB 23;  Length 560;
Best Local Similarity   57.7%;  Pred. No. 8.2;
Matches   71;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy   875 atgcactttaatacggcttcttgtgctatattatttatttcatggctgtatctagt 934
          |||  | ||| |  ||| | |||  || ||| |||  || ||   || | || ||||
Db   125 ATGAAGTTCCAATAATTTTTTTTTTTTCCTTCTTTTCGTTGTTGTTTTTAGAT 66

Qy   935 ttgcatttaagatttagtaccttgtcagattaaagaaaacgaaagttaaattaaaaa 994
          |||||| ||| |    ||    ||||  |||||||||||||| ||| |||||| |||||
Db    65 AAAAATTTGAGAAAAAAAATGAAAAAATTTAATAAAAGAAGAAAATGAAAAAAAAA 6

Qy   995 aaa 997
          |||
Db     5 AAA 3

RESULT  14
LOCUS       C91220                    574 bp    mRNA              EST        15-APR-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSJ834, mRNA sequence.
ACCESSION   C91220
NID         g3060586
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 574)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 574)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..574
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="slug"
BASE COUNT      258 a     76 c     63 g    176 t      1 others
ORIGIN
```

```
Query Match              4.0%;  Score 39.8;  DB 28;  Length 574;
Best Local Similarity   60.7%;  Pred. No. 8.2;
Matches   65;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy   227 acttcaaaaacaaagcatcatggcaagacagatggagtacttggagtacttgaataacaacaatggcg 286
          |  |||||||||| |   ||||| |   ||||||||||||| ||||| |||||||||||||||||||||
Db   386 AATTGAAAAACAATATATATAATAGAAAAGCAACAACAACAACAACAACAACAACAACAACA 445

Qy   287 acaataacaacaataatgttacaagtcattacgagataattat 333
          |||| ||||||||||| | | ||||||| || || |||||||||
Db   446 ACAACAACAACAACAACAACAACAACAACAACAACAACAAAAAATAATTAT 492

RESULT  15
LOCUS       C22974                    546 bp    mRNA              EST        27-MAR-1997
DEFINITION  Dictyostelium discoideum gamete cDNA, clone FC-AL20, mRNA sequence.
ACCESSION   C22974
NID         g1922123
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:KAX3) Sexually mature Gamete cDNA
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 13:56:00 ; Search time 822.9 Seconds
                 (without alignments)
                 2686.748 Million cell updates/sec

Title: US-09-156-580-1_COPY_190_807
Perfect score: 618
Sequence: 1 atggcaggcatgatagaaa.........aacttcgactgggatccact 618
Scoring table: IDENTITY_NUC
Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl:*
    1: gb_bal:*
    2: gb_ba2:*
    3: gb_in:*
    4: gb_om:*
    5: gb_ov:*
    6: gb_pat:*
    7: gb_ph:*
    8: gb_pl1:*
    9: gb_pl2:*
   10: gb_pr1:*
   11: gb_pr2:*
   12: gb_pr3:*
   13: gb_ro:*
   14: gb_st:*
   15: gb_sy:*
   16: gb_un:*
   17: gb_vi:*
   18: gb_htg:*
   19: gb_ba:*
   20: em_fun:*
   21: em_hum1:*
   22: em_hum2:*
   23: em_in:*
   24: em_om:*
   25: em_or:*
   26: em_ov:*
   27: em_pat:*
   28: em_ph:*
   29: em_pl:*
   30: em_ro:*
   31: em_sy:*
   32: em_un:*
   33: em_vi:*
   34: em_htg:*
   35: em_sts:*
   36: gb_bal:*
   37: gb_ba2:*
   38: gb_pl1:*
   39: gb_pl2:*
   40: gb_pr1:*
   41: gb_pr2:*
   42: gb_pr3:*
   43: gb_sts:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Str	Score	Query Match	Length	DB	ID	Description
1		89	14.4	1071	9	ATU38946	U38946 Arabidopsis
2	c	89	14.4	1071	39	ATU38946	U38946 Arabidopsis
3	c	76.2	12.3	99687	18	AC005956	AC005956 *** SEQUE
4	c	70.4	11.4	93845	39	ATAC004684	AC004684 Arabidops
5		70.4	11.4	93845	39	ATAC004684	AC004684 Arabidops
6		49	7.9	5931	3	DDU32174	U32174 Dictyosteli
7		46.8	7.6	1200	8	MISCVA	V00705 Yeast mitoc
8		46.8	7.6	1200	8	YSCMTVAR1	J01525 Yeast (S.ce
9		46.8	7.6	1200	38	MISCVA	V00705 Yeast mitoc
10		46.2	7.5	1200	8	YSCMTVAR1	J01525 Yeast (S.ce
11		46.2	7.5	1083	38	PET2FDB2	D26084 Petunia mRN
12		44.6	7.5	1083	8	PET2FDB2	D26084 Petunia mRN
13		44.6	7.2	3041	38	PET2FDB1	D26083 Petunia zin
14		44.2	7.2	3041	3	PET2FDB1	D26083 Petunia zin
15		44.2	7.2	12668	3	AF079445	AF079445 Dictyoste
16		43.8	7.1	1453	3	DD1039PPR1	AF018638 Plasmodiu
17		43.8	7.1	7467	38	PFSC04099	AL010282 Plasmodiu
18		43.8	7.1	5100	38	YSCPHO81	D13228 Yeast PHO81
19		42.2	6.8	5100	8	YSCPHO81	D13228 Yeast PHO81
20		42.2	6.8	3478	38	DDIADCYG	M87278 Dictyosteli
21		42	6.8	3478	3	DDIADCYG	M87278 Dictyosteli
22		41.8	6.8	4374	3	AF045453	AF045453 Dictyoste
23		41.4	6.7	235	43	PFPPF297	X53020 P.faciparu
24		41.2	6.7	165	8	G37866	G37866 emaa2 Plasm
25		41.2	6.7	6265	8	PFSC03080	AL010153 Plasmodiu
26		41.2	6.7	2107	38	DDIADCYA03	L05498 Dictyosteli
27		41.2	6.7	2107	38	DDIADCYA03	L05498 Dictyosteli
28		41.2	6.7	221	18	G37927	G37927 c3m61 Plasm
29		41	6.6	93419	41	AC002415	AC002415 *** SEQUE
30	c	41	6.6	231	43	PC064653	U64653 Paracharter
31		40.8	6.6	7133	3	AF024654	AF024654 Dictyoste
32		40.8	6.6	3887	3	AF029726	AF029726 Dictyoste
33		40.8	6.6	2239	3	DDU07817	U07817 Dictyosteli
34		40.8	6.6	6115	9	DDU20432	U20432 Dictyosteli
35		40.8	6.6	1191	39	SCE010480	AJ010480 Saccharom
36		40.8	6.6	1191	6	SCE010480	AJ010480 Saccharom
37		40.6	6.6	5108	24	BBU44914	U44914 Borrelia bu
38		40.6	6.6	1690	37	I52144	I52144 Sequence 24
39		40.6	6.6	1700	38	S79359	S79359 arabinogala
40		40.6	6.6	5108	2	BBU44914	U44914 Borrelia bu
41		40.6	6.6	1700	5	S79359	S79359 arabinogala
42		40.4	6.5	60232	8	AE001272	AE001272 Lactococc
43	c	40.4	6.5	41076	18	CER02D5	Z78015 Caenorhabdi
44		40.4	6.5	1640	8	AB000453	AB000453 Petunia h
45	c	40.4	6.5	232854	18	CEY40H4	ALO22573 Caenorhab

ALIGNMENTS

RESULT   1
ATU38946
LOCUS   ATU38946   1071 bp   DNA   PLN   29-NOV-1995
DEFINITION Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.
ACCESSION   U38946
NID   g1079668
KEYWORDS   .
SOURCE   thale cress.
  ORGANISM Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyte/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 1071)
  AUTHORS   Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
  TITLE   Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries
  JOURNAL   Nature 378 (6553), 199-203 (1995)
  MEDLINE   96069740
REFERENCE   2 (bases 1 to 1071)
  AUTHORS   Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech,
            Pasadena, CA 91125, USA
FEATURES   Location/Qualifiers

```
source          1..1071
                /organism="Arabidopsis thaliana"
                /strain="Landsberg erecta"
                /db_xref="taxon:3702"
                /chromosome="3"
5'UTR           join(1..68,175..202)
gene            /gene="sup"
                1..1071
                /gene="sup"
mRNA            join(1..68,175..1071)
                /gene="sup"
intron          69..174
                /gene="sup"
CDS             203..817
                /gene="sup"
                /standard_name="SUPERMAN"
                /note="Allele: Ler wild type;  C2H2 zinc finger protein"
                /codon_start=1
                /function="flower-specific gene controlling the boundary
                of the stamen and carpel whorls"
                /product="SUPERMAN"
                /db_xref="PID:g1079669"
                /translation="MERSNSIELRNSFYGRARTSPWSYGDYDNCQQDHDYLLGFSWPP
                RSYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSTPSPPYPNPNYSYSTM
                ANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSSLLVEAGE
                ATRFTSKDACKILRNDEIISLELEIGLINESEQDLDLELRLGFA"
3'UTR           818..1071
                /gene="sup"
BASE COUNT      326 a    226 c    183 g    336 t
ORIGIN

Query Match           14.4%;     Score 89;  DB 9;  Length 1071;
Best Local Similarity 68.9%;    Pred. No. 9.7e-10;
Matches 122;  Conservative    0;  Mismatches 55;  Indels  0;  Gaps  0;

Qy   124 agctcattacgagataattatggaaatgaagatcattacttggtggactattccttgg 183
         ||||  |  ||| | || | | |  ||| |||||| |||||| | || ||| |||
Db   269 AGCTATGGAGAGATTATGATAATTGCCAACAGGATCATGATTATCTTCTAGGGTTTTCATGG 328

Qy   184 cctccaagatcttatacatgtagcttttgtaaaagggaatttagatctgctcaagctctt 243
         || |||||| ||  || |||| | || || |  ||||||| |||||| |||||| |||
Db   329 CCACCAAGATCCTACACTTGCAGCTTCTGCAAAAGGGAATTCAGATCGCTCAAGCACTT 388

Qy   244 ggtggacacatgaatgttcatagagagatagagccatttgagacaatcaccaccct 300
         ||||| || || | || ||||||||||| || | ||  || ||  |  || | ||
Db   389 GGTGGCCACATGAATGTTCACAGAGACAGAGAGAGACTCAGATTACAACAGTCT 445


RESULT 2
ATU38946
LOCUS       ATU38946     1071 bp    DNA             PLN       29-NOV-1995
DEFINITION  Arabidopsis thaliana SUPERMAN (sup) gene, complete cds.
ACCESSION   U38946
NID         g1079668
KEYWORDS    .
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; Magnoliophyta;
            Magnoliopsida; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1071)
AUTHORS     Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
TITLE       Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries
JOURNAL     Nature 378 (6553), 199-203 (1995)
MEDLINE     96069740
REFERENCE   2  (bases 1 to 1071)
AUTHORS     Sakai,H., Medrano,L.J. and Meyerowitz,E.M.
TITLE       Direct Submission
JOURNAL     Submitted (19-OCT-1995) Hajime Sakai, Biology, 156-29, Caltech,
            Pasadena, CA 91125, USA
FEATURES             Location/Qualifiers
source          1..1071
```

```
source          
                /organism="Arabidopsis thaliana"
                /strain="Landsberg erecta"
                /db_xref="taxon:3702"
                /chromosome="3"
5'UTR           join(1..68,175..202)
                /gene="sup"
gene            1..1071
                /gene="sup"
mRNA            join(1..68,175..1071)
                /gene="sup"
intron          69..174
                /gene="sup"
CDS             203..817
                /gene="sup"
                /standard_name="SUPERMAN"
                /note="Allele: Ler wild type;  C2H2 zinc finger protein"
                /codon_start=1
                /function="flower-specific gene controlling the boundary
                of the stamen and carpel whorls"
                /product="SUPERMAN"
                /db_xref="PID:g1079669"
                /translation="MERSNSIELRNSFYGRARTSPWSYGDYDNCQQDHDYLLGFSWPP
                RSYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSTPSPPYPNPNYSYSTM
                ANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSSLLVEAGE
                ATRFTSKDACKILRNDEIISLELEIGLINESEQDLDLELRLGFA"
3'UTR           818..1071
                /gene="sup"
BASE COUNT      326 a    226 c    183 g    336 t
ORIGIN

Query Match           14.4%;     Score 89;  DB 39;  Length 1071;
Best Local Similarity 68.9%;    Pred. No. 9.7e-10;
Matches 122;  Conservative    0;  Mismatches 55;  Indels  0;  Gaps  0;

Qy   124 agctcattacgagataattatggaaatgaagatcattacttggtggactattccttgg 183
         ||||  |  ||| | || | | |  ||| |||||| |||||| | || ||| |||
Db   269 AGCTATGGAGAGATTATGATAATTGCCAACAGGATCATGATTATCTTCTAGGGTTTTCATGG 328

Qy   184 cctccaagatcttatacatgtagcttttgtaaaagggaatttagatctgctcaagctctt 243
         || |||||| ||  || |||| | || || |  ||||||| |||||| |||||| |||
Db   329 CCACCAAGATCCTACACTTGCAGCTTCTGCAAAAGGGAATTCAGATCGCTCAAGCACTT 388

Qy   244 ggtggacacatgaatgttcatagagagatagagccatttgagacaatcaccaccct 300
         ||||| || || | || ||||||||||| || | ||  || ||  |  || | ||
Db   389 GGTGGCCACATGAATGTTCACAGAGACAGAGAGAGACTCAGATTACAACAGTCT 445


RESULT 3
AC005956/c
LOCUS       AC005956     99687 bp    DNA             HTG       04-DEC-1998
DEFINITION  *** SEQUENCING IN PROGRESS *** Arabidopsis thaliana 'MITSUI' P1
            'MHK10' genomic sequence near marker 'EG05C12'; HTGS phase 2, 1
            ordered pieces.
ACCESSION   AC005956
NID         g3962504
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
            Magnoliophyta; eudicotyledons; Rosidae; Capparales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 99687)
AUTHORS     Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M., Shen,M.,
            Ronning,C.M., Rounsley,S.D., Fraser,C.M., Somerville,C.R. and
            Venter,J.C.
TITLE       Arabidopsis thaliana 'MITSUI' BAC 'MHK10' genomic sequence near
            marker 'EG05C12'
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 99687)
AUTHORS     Rounsley,S.D. and Lin,X.
TITLE       Direct Submission
```

```
JOURNAL    Submitted (10-NOV-1998) The Institute for Genomic Research, 9712
           Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
COMMENT    On Dec 4, 1998 this sequence version replaced gi:3894154.
           ***
           *** WARNING: Phase 2 High Throughput Genome Sequence ***
           ***
           * This sequence is unfinished. It consists of 1 contigs for
           * which the order is known. The lengths of the gaps have been
           * estimated by the submitter but are not known exactly. When
           * sequencing is complete, the sequence data presented in this
           * record will be replaced by a single finished sequence
           * with the same accession number.
           *
           1    99687: contig of 99687 bp in length.
FEATURES   Location/Qualifiers
   source  1..99687
           /organism="Arabidopsis thaliana"
           /db_xref="taxon:3702"
           /clone="MHK10"
BASE COUNT   30477 a  18393 c  19273 g  31544 t
ORIGIN

Query Match          12.3%;   Score 76.2;  DB 18;   Length 99687;
Best Local Similarity  72.3%;  Pred. No. 2.4e-07;
Matches  99;  Conservative  0;  Mismatches  38;  Indels  0;  Gaps  0;

Qy    166 ggtggcatctcctggcctccaagatcttatacatgtagctttgtaaaaggaattt 225
          ||||||||||||||||| |||| |||||||||||||||||||| ||||||| |||||
Db  45805 GCTGGGCTGATCCCCATGGCCTCCCAAGAATTACACTTGCAGCTTCTGCAGGAGAGAGTTC 45746

Qy    226 agatctgtcaagctcttggtggacacatgaatgttcatagaagagatagagccatttg 285
          |||||||||||| ||||| |||||||||||||||||||||||||||| |||| |
Db  45745 AGATCTGCTCAAGCACTTGGAGGCCCATGAATGTTCATAGAAGAGACCAGCAAAAACTC 45686

Qy    286 agacaatcaccacctag 302
          ||||| ||| ||||
Db  45685 AGGCAGATCCCCTTCTTG 45669

RESULT  4
ATAC004684/C
LOCUS       ATAC004684   93845 bp    DNA            PLN       17-JUN-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence,
            complete sequence.
ACCESSION   AC004684
NID         g3236234
KEYWORDS    HTG
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,
            Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jun 17, 1998 this sequence version replaced gi:3172168.
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr,
```

```
            Rockville, MD 20850,
            USA
            e-mail: rounsley@tigr.org
            BAC clone F13M22 is from Arabidopsis chromosome II and is near the
            molecular marker ve018.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without similarity to a significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as genes but have predicted
            exons by GRAIL are annotated as misc features.
FEATURES     Location/Qualifiers
   source    1..93845
             /organism="Arabidopsis thaliana"
             /cultivar="Columbia"
             /db_xref="taxon:3702"
             /chromosome="II"
             /map="ve018"
             /clone="F13M22"
   misc_feature  complement(10..29)
   misc_feature  168..249
             /note="exon predicted by xgrail, quality marginal"
   mRNA     <464..>619
             /note="exon predicted by xgrail, quality good_shadowexon"
   gene     <464..>619
             /gene="F13M22.1"
   CDS      <464..619
             /gene="F13M22.1"
             /note="predicted by genscan"
             /gene="F13M22.1"
             /note="hypothetical protein, 5' partial"
             /codon_start=1
             /db_xref="PID:g3236260"
             /translation="SRGDVRVEQRRTHKDGEDKRLLDPATRRKLETNAIIASGTKGF
             EIGNGNL"
   mRNA     complement(join(835..1212,1373..1513,1600..1693,
             1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
             2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
             4167..4353,4486..4549,4633..4842,4938..5006,5088..>5654))
             /gene="F13M22.2"
   gene     complement(835..>5654)
             /gene="F13M22.2"
   CDS      complement(join(1099..1212,1373..1513,1600..1693,
             1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
             2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
             4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
             /gene="F13M22.2"
             /note="unknown protein"
             /codon_start=1
             /db_xref="PID:g3236235"
             /translation="MPKEENGQLKRDRLDDDTDEGNKGDHFPSKKQAKEASNDDITSE
             ISNPVASPVESTSLFRDVSSOPVKSGLVECSGSDFGSEETVSDDASVVGSSOTEQSSD
             VLPSRFVLEIPKHLSSTGITKITFKLSKPKKEFDDLPLIKDHTWDAGVVKMPKKIVS
             LSYPSNVKLLETGILEGARVKYISTPPVRQLLGIIHSGGYLCGCTCNFSKVLSAYE
             FEQHAGAKTRHPNNHIFLENRRAVVNIVQELKTAPRVVLEEVIRNVAGSALNEEGLRA
             WKGDGQVSNRVIACLTGIILRIIPLSVNALLSYHFVLLSETDSCESCSYLGPGLDES
```

```
                QSLTPCSVENHYFSEKTYAKDTLDEPKRIAKKLTSHVSGTGCHKKVSEGSNRKRDNDL
                HRLLFMPNGLPDGTELAYYVKTQISPSQFEAHAGMAARRQPYRHIFISSGLSLHDIAM
                SLANGHVITTGDSDDMCSICGDGGDLLLCAGCPQAFHTACLKFQSMPEGTWYCSSCND
                GPISSKKATTTDPSGNARPIVIRLSRVVKAPESDIGGCVFCRSHDFSIGKFDDRTVIL
                CDQCEKEYHVGCLRENGFCDLKEIPQEKHFCCSNCSRIHTAVQNSVSCGPQTLPTPLL
                DMICRKDREKGIFTDIGDTVEWRILSGKSRYPEHLPLLSRAAVIFRECFDPIVAKSGR
                DLIPVMVYGRNISGQEFGGMYCLVLIVNSLVVSAALLRIFGQEVAELPIVATSREYQG
                RGYFQGLYACVENLLSSLNVENLVLPAAEEAESIWTKKFGFTKMSDQQLQEYQKEVQL
                TIFKGTSMLEKKVPKATTGLSESTTLI"
repeat_region   6010. .6135
                /rpt_family="(GGAA)n"
misc_feature    6023. .6202
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   6179. .6236
                /rpt_family="POLY_A"
mRNA            join(<6647. .6702,7161. .7304,7322. .>7499)
                /gene="F13M22.3"
gene            <6647. .>7499
                /gene="F13M22.3"
                /note="predicted by genscan"
CDS             join(6647. .6702,7161. .7304,7322. .7499)
                /gene="F13M22.3"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236236"
                /translation="MTTWKKVREENKSFQFLFKFAWRIPHFVVGATWTLFLILTVAII
                SLAPEFAFVSAIFPSSSEVFSRRHIGSAGSSLGGLVFTGESFPEVEDGPRCAAGFCSN
                CCGSFRRRRKDYGPLGGRRSPLI"
mRNA            complement(join(<7761. .7931,8019. .8065,8154. .8238,
                8323. .8401,8494. .8647,8824. .8903,9001. .9232,9769. .>9886))
                /gene="F13M22.4"
gene            complement(<7761. .>9886)
                /gene="F13M22.4"
CDS             complement(join(7761. .7931,8019. .8065,8154. .8238,
                8323. .8401,8494. .8647,8824. .8903,9001. .9232,9769. .>9886))
                /gene="F13M22.4"
                /product="putative ribotol dehydrogenase"
                /codon_start=1
                /db_xref="PID:g3236237"
                /translation="MGIYGVMTGKKGKSGFGSASTAEDVTQAIDASHLTAIITGGTSG
                IGLEAARVLAMRGAHVIIAARNPKAANESKEMLQMNPNARVDYLQIDVSSIKSVRSF
                VDQFLALNVPLNLIINNAGVMFCPFKLTEDGISQFATNHIGHFLLTNLLLDKMKSTA
                RESSVQGRIGSIIAHTTYSGLTSRVPFIKRHDRTFKLTDSASEHCPPGEFVREIPLSNRIPEDAN
                RLQEEGVNITINSVHPGLVTTNLFRYSGFSMKVFRAMTFLFWKNIPQGAATTCYVALH
                PDLEGVTGKYFGDCNIVAPSKFATNNSLADKLWDFSVFLIDSISK"
repeat_region   8658. .8884
                /rpt_family="AT_rich"
repeat_region   complement(9842. .9915)
                /rpt_family="(GA)n"
misc_feature    10503. .10670
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   10725. .10746
                /rpt_family="AT_rich"
mRNA            complement(join(11416. .11526,11608. .11616,12551. .12628,
                12737. .13604))
                /gene="F13M22.5"
gene            complement(11416. .13604)
                /gene="F13M22.5"
CDS             complement(join(11610. .12551. .12628,12737. .13522))
                /gene="F13M22.5"
                /product="putative ARF1 GTPase activating protein"
                /codon_start=1
                /db_xref="PID:g3236238"
                /translation="MAAARRLRFLQSQPENKVCVDCSQKNPQWASISYGIFMCLECSG
                KHRGLGVHISFVRSVTMDSWSEIQIKKMDAGGNERLNNFLAQYGISKETDIISKYNSN
                AASVYDRIQALAEGRQWRDPPIVKESVGGLMNKKPPLSQGGRDSGNGGWNDWG
```

```
                TGGGFL"
repeat_region   11697. .11757
                /rpt_family="(CAT)n"
misc_feature    13716. .13952
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
misc_feature    complement(14172. .14309)
                /note="exon predicted by xgrail, quality marginal"
mRNA            join(14707. .14911,15566. .15765,15858. .15964,16057. .16189,
                16311. .16602,16755. .17294)
                /gene="F13M22.6"
gene            14707. .17294
                /gene="F13M22.6"
                /note="identical to GB:U40269"
CDS             join(15634. .15765,15858. .15964,16057. .16189,16311. .16602,
                16755. .17182)
                /gene="F13M22.6"
                /codon_start=1
                /product="origin recognition complex protein"
                /db_xref="PID:g3236239"
                /translation="MEDIENIEEDEYGFSRNYFLAKELGGASKRSAHKLSDIHIVDEQ
                ELRETASTIEMKHSKEISELMSDYKTMYSKWVFELRCGFGLLMYGFGSKKALVEDFAS
                ASLTDYSVVVINGYLPSVNLQVLALALLSELLKCKRKSSGSLSKGQETFPSRSMD
                DILSFLHGPQSGDKDCFICVVHNIDGPALRDPESQQTLARLSSCSHIRLVASIDHVN
                APLLWDKKMVHKQFNWLWHVPTFAPYNVEGVFFPLVLAQGSTAQTAKTAAIVLQSLT
                PNGQNVFKILAEYOLSHPDEDGMPTDDLYSASRERFFVSSQVTLNSHLTEFKDHELVK
                TKRNSDGQECLNIPLTSDAIRQLLDLNQ"
mRNA            complement(<17532. .>19016)
                /gene="F13M22.7"
gene            complement(<17532. .>19016)
                /gene="F13M22.7"
CDS             complement(17532. .19016)
                /gene="F13M22.7"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236240"
                /translation="MENHHPSTLLSMDSSASSHEELDLEMNNNRQSLLSGPPDINLPL
                SAERSPPPPWNLDACDILDVGLGSQAYETENYMSVVPKVGRCAKRVDSIWGAWFFF
                SFVFKPALNEKSAKIVRDSNGISGFDKSDLKLDVFLVQHDMENMYMWVFKERPENAL
                GKMQLRSYMNGHSRQGDRLFPFSVEKGFVRSHRMQRKHYRGLSNPQCVHGIELVPLPN
                LTCLDEEERKRWMELTGRDLNFTIPPEASDFGSWRNLPNTDFELERPTPSLKNPTANH
                SKKLLNGSGLNLSTQPSNHSNGEATDLSPSSHKKRKDLFSNGIHEEECCLTVNPQQPPV
                IEAHQNELPTWSNEFTGAMKNVIGPVTAKTIYEDEEGYLIIISLPFVDLNSVKVSWR
                NTLTHGIIKVSCLSTSRVPFIKRHDRTFKLTDSASEHCPPGEFVREIPLSNRIPEDAN
                IEAYYDGPGSVLEILVPKLRAGPEEHEVRVCLRPNLGGNDLMLT"
misc_feature    complement(19045. .19091)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    19455. .19545
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
mRNA            join(<20811. .21507,21667. .21805,22028. .22434,22931. .23006,
                23118. .23233,23331. .>23788)
                /gene="F13M22.8"
gene            <20811. .>23788
                /gene="F13M22.8"
CDS             join(20811. .21507,21667. .21805,22028. .22434,22931. .23006,
                23118. .23233,23331. .23788)
                /gene="F13M22.8"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236261"
                /translation="MFNNTTTSFGSGPGIVVVPTPATTVPTDFPGTTITSNSTFIII
                GPPPPFPAPPRSIDLTPLKLIFVVIAFVAVPALVYALFFNGPCSSSRRNSSSSRTSSS
                SDDTPHATVDTPPITETTVTSESGGKFHKDTHSKEIGNECSVCLMVFTDSDELRQLSE
                CKHAFHVLCIETWLKDHPNCPICRTDVSVRQTCAPNVPVNVNRSGNRRVSAT
                SRDDDWRQGLPDASTTVSGVSRDVKSDVDSQQAHPTVIEDLILGRLLIHTRLPIRTL
                SFCSRKPSDSSSGGLAPNRNLATKSTIPRFAYLVTGTKGDGKRVKRLLKAIHHPRNYL
                LHLDLEASDEERMELAKYVRSEKKFENVMVMGLADLVTEKGPTMLASTLHGVALLK
                KAKDWDWFINLSASDYPLMPQDDILHIFSYLPRYLNFIEHTSNIGWKENQRARPIIID
                PGFYHLKKSGVFWAKERRSLPASFKLFMGSTSVALTRPFLEFCIWGWDNLPRTLLMYY
                TNFLLSSEGYFQTVVCNNHDLHYTKWDPLQQRTLNVTVENFRDMVQSGA
                PFAREFREDDLVLDKIDIELLGQTDGLEKTPDVVKPTVSWRKLEKLMVRLLDHENF
                RAKQCK"
```

```
repeat_region   complement(20872..20971)
                /rpt_family="(TGG)n"
repeat_region   complement(21518..21543)
                /rpt_family="AT_rich"
misc_feature    complement(24200..24291)
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   complement(25078..25159)
                /rpt_family="(TA)n"
mRNA            join(<25358..25408,25622..>26563)
                /gene="F13M22.9"
gene            <25358..>26563
                /gene="F13M22.9"
CDS             join(25358..25408,25622..26563)
                /gene="F13M22.9"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236241"
                /translation="MVFSSIQAYLDSSNWQQAPPSNYNHDGTGASANGGHVLRPQLQP
                QQPQOQPHPNGSGGGGGGGSIRAGSMVDRARQANVALPEAALKCPRCESTNTKFC
                YFNNYSLTQPRHFCKTCRRYWTRGGALRNVPVGGCRRNRRTKSNSNNNNSTATSNN
                TSFSSGNASTISTILSSHYGGNQESILSQILSPARLMNPTYNHLGDLTSNTKTDNNMS
                LLNYGGLSQDLRSIHMGASGSGSLMSCVDEWRSASYHQQSSMGGGNLEDSSNPNPSANG
                FYSFESPRITSASISSALASQFSSVKVEDNPYKWVNVNGNCSSWNDLSAFGSSR"
repeat_region   25691..25728
                /rpt_family="(CAG)n"
repeat_region   26002..26054
                /rpt_family="(CAA)n"
misc_feature    complement(27038..27122)
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   28179..28208
                /rpt_family="AT_rich"
misc_feature    complement(28259..28328)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   complement(28897..28980)
                /rpt_family="(TA)n"
repeat_region   29086..29133
                /rpt_family="AT_rich"
misc_feature    complement(29945..30023)
                /note="exon predicted by xgrail, quality good"
mRNA            complement(join(<30476..30515,30659..30858,31168..>31269))
                /gene="F13M22.10"
gene            complement(<30476..>31269)
                /gene="F13M22.10"
CDS             complement(join(30476..30515,30659..30858,31168..>31269))
                /gene="F13M22.10"
                /codon_start=1
                /product="putative ribosomal protein L36"
                /db_xref="PID:g3236242"
                /translation="MTTPAVKTGLFVGLNKGHVVTRRELAPPNSRKGKTSKRTIFIR
                KLIREVAGMAPYEKRITELLKVGKDKRALKVAKRKLGTHKRAKRKREEMSSVLRKMRS
                LGGAAAEKKM"
misc_feature    complement(31377..31401)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    complement(32006..32036)
                /note="exon predicted by genscan"
repeat_region   32044..32161
                /rpt_family="POLY_A"
mRNA            complement(join(<32935..33310,33630..>33739))
                /gene="F13M22.11"
gene            complement(<32935..>33739)
                /note="predicted by genscan"
CDS             complement(join(32935..33310,33630..33739))
                /gene="F13M22.11"
                /codon_start=1
                /note="hypothetical protein"
                /db_xref="PID:g3236243"
                /translation="MEIGNVKADVNLVLNLREAERHRRYFDYFIDEIIWLSFLGMSET
                SLKEEVTLGGVKETREEECSDDKLVLDICLKRKRLDEEEEVVEEENDGFKTPTRP
                ENRIPIVRECPPAPMKRSSEMFRGRTMYCRRRLSFLPEDDVNSFITDLQWRTTTMTIK
                K"
repeat_region   complement(33140..33196)
                /rpt_family="(GAA)n"
repeat_region   33357..33398
                /rpt_family="(GAA)n"
repeat_region   33469..33489
                /rpt_family="AT_rich"
misc_feature    complement(34248..34296)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    35206..35258
                /note="exon predicted by xgrail, quality good"
repeat_region   complement(35273..35341)
                /rpt_family="(GGA)n"
mRNA            join(35801..35886,36023..36416,36517..37130,37242..37322)
                /gene="F13M22.12"
gene            35801..37322
                /note="identical to GB:U29480"
                /gene="F13M22.12"
CDS             join(35827..35886,36023..36416,36517..37130,37242..37307)
                /gene="F13M22.12"
                /codon_start=1
                /product="actin 3 protein"
                /db_xref="PID:g3236244"
                /translation="MADGEDIQPLVCDNGTGMVKAGFAGDDAPRAVFPSIVGRPRHTG
                VMVGMGQKDAYVGDEAQSKRGILTLKYPIEHGIVNNWDDMEKIWHHTFYNELRVAPEE
                HPILLTEAPLNPKANREKMTQIMFETFNAPAMYVAIQAVLSLYASGRTTGIVLDSGDG
                VSHTVPIYEGYALPHAILRLDLAGRDLTDALMKILTERGYSFTTTAEREIVRDIKEKL
                CYIALDYEQELETAKTSSSVEKNYELPDGQVITIGSERFRCPEVLYQPSMIGMENAGI
                HETTYNSIMKCDVDIRKDLYGNIVLSGGTTMFPGIADRMSKEITALAPSSMKIKVVAP
                PERKYSVWIGGSILASLSTFQQMWIAKAEYDESGPSIVHRKCF"
repeat_region   37700..37817
                /rpt_family="(CAAA)n"
mRNA            complement(<37864..>38967)
                /gene="F13M22.13"
gene            complement(<37864..>38967)
                /gene="F13M22.13"
CDS             complement(37864..38967)
                /gene="F13M22.13"
                /codon_start=1
                /product="myb-related protein"
                /db_xref="PID:g3236245"
                /translation="MKERQRWSGEEDALLRAYVRQFGPREWHLVSERMNKPLNRDAKS
                CLERWKNYLKPGIKKGSLTEEEQRLVIRLQEHNVVPAAAAATVVMANSN
                FKEKQQREEKESNKRVEPIDESKYDRILESFAEKLVKERSNVVPAAAAATVVMANSN
                GGFLHSEQOVQPPNPVIPPWLATSNNGNNVVARPPSVTLTLSPSTVAAAPQPPIPWL
                QOQOPERAEENGPGGLVLGSMPSCSGSSESVFLSELVECCRELEEGHRAWADHKKEAA
                WRLRRLELQLESEKTCRQREKMEIEAKMKALREQKNAMEKIEGEYREQLVGLRRDA
                EAKDQKLADQWTSRHIRLTKFLEQQMGCRLDRP"
repeat_region   complement(39531..39561)
                /rpt_family="AT_rich"
repeat_region   complement(40737..40818)
                /rpt_family="(TAA)n"
repeat_region   complement(41089..41112)
                /rpt_family="POLY_A"
repeat_region   complement(41190..41213)
                /rpt_family="POLY_A"
misc_feature    complement(41283..41310)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   complement(41558..41602)
                /rpt_family="AT_rich"
repeat_region   complement(41966..42068)
                /rpt_family="(GAAA)n"
repeat_region   complement(42281..42325)
                /rpt_family="AT_rich"
misc_feature    42889..42968
                /note="exon predicted by xgrail, quality excellent"
misc_feature    43132..43154
                /note="exon predicted by xgrail, quality good"
repeat_region   43371..43479
                /rpt_family="AT_rich"
```

```
mRNA            /rpt_family="(TAAAA)n"
                complement(join(43750. .44452,45240. .45558,45719. .45963))
gene            /gene="F13M22.14"
                complement(43750. .45963)
CDS             /gene="F13M22.14"
                complement(join(44143. .44452,45240. .45558,45719. .45878))
                /gene="F13M22.14"
                /codon_start=1
                /product="putative expansin protein"
                /db_xref="PID:g3236246"
                /translation="MTATAFRVGLMWLAVTASFLLTATNAKIPGVVSGGPWQNAHATFY
                GGSDASGTMGGACGYGNLYSQGYGVNTAALSTALFNNGFSCGACFEIKCTDDPRWCVP
                GNPSLVTATNFCPPNFAQPSDDGWCNPPREHFDLAMPMFLKIGLYRAGIVPVSYRR
                VPCRKIGGIRFTVNGFRYFNLVLVTNVAGAGDINGYSVKGSKTDWVRMSRNWGQNWQS
                NAVLIGQSLSFRVTASDRRSSTSWNVAPATWQFGQTFSGKNFRV"
repeat_region   44795. .44903
                /rpt_family="AT_rich"
repeat_region   46287. .46312
                /rpt_family="AT_rich"
repeat_region   complement(47301. .47444)
misc_feature    /note="exon predicted by xgrail, quality
                excellent_shadowexon"
                48559. .48589
repeat_region   /rpt_family="AT_rich"
mRNA            <48689. >50845
                /gene="F13M22.15"
gene            <48689. >50845
                /gene="F13M22.15"
CDS             48689. .50845
                /gene="F13M22.15"
                /codon_start=1
                /product="SCARECROW-like protein"
                /db_xref="PID:g3236247"
                /translation="MITEPSLTGISGMVNNRLSGLPDQPSSHSFTPVTLYDGFNYNL
                SSDHINTVVAAPENSVFIREEEEEDPADDFDSAVLGYISQMLNEEDMDDKVCMLQ
                ESLDLEAAERSLYEAIGKKYPPSPERNLAFAERNSENLDRVVPGNYTGDCIGFGNGG
                IKPLSSGFTLDFRNPQSCSSILSVPQSNGLITIYGDGIDESSKNNRENHQSVWLFRRE
                IEEANRFNPENELIVNFRENRVSXARKNSSRDEICVEEERSSKLPAVFGEDILRSD
                VVDKILVHVPGGESMKEFNALRDVLKKGVEKKKASDAQGGKRRARGRGRGRGGGG
                QNGKKEVDLRSLLIHCAQAVAADDRRCAGQLLRQIRLHSTPFGDGNQRLAHCFANGL
                EARLAGTGSQIYKGIVSKPRSAAAVLKAHQLFLACCPFRKLSYFTINKTIRDLVQNSQ
                RVHVIDFGILYGFQWPTLIHRFSMYGSPKVRITGIEFPQPGFRPAQRVEETGQRLAAY
                AKLFGVPFEYKAIAKKWDAIQLEDLDIRDEITVVNCLYRAENLHDESVKVESCRDTV
                LNLIGIKINPDALVFVTFVGAYNAPFFVTFREALFHFSSSIFDMLETIVPREDEERMFL
                EMEVPGREALNVIACEGWERVERPETYKQWHVRAMRSGLVQVPFDPSIMKTSLHKVHT
                FYHKDFVIDQDNRWLLQGWKGRTVMALSVKKPESKA"
repeat_region   49591. .49709
                /rpt_family="(GAA)n"
repeat_region   49635. .49689
                /rpt_family="(GGA)n"
misc_feature    51112. .51155
misc_feature    complement(51301. .51387)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
mRNA            complement(join(51428. .51666,51782. .51905,51988. .52049,
                52131. .52264,52306. .52342,52427. .52698,52779. .>53043))
                /gene="F13M22.16"
gene            complement(51428. .>53043)
                /gene="F13M22.16"
                /note="unknown protein"
CDS             complement(join(51547. .51666,51782. .51905,51988. .52049,
                52131. .52264,52306. .52342,52427. .52698,52779. .53043))
                /gene="F13M22.16"
                /codon_start=1
                /db_xref="PID:g3236248"
                /translation="MAMMTTTTTTFFHPLLPANTYKSGAVASSFVSVPRSSSLQFRSL
                VSDSTSICGPSKFTGKNRRYSVTVSAAATTEPLTVLVTGAGGRTGQIVYKKLKERSEQ
                FVARGLVRTKESKEKINGEDEVFIGDIRDTASIAPAVEGIDALVILTSAVPQMKPGFD
                PSKGGRPEFFFDDGAYPEQVWIGQKNQIDAGVDDDGLSFVVTAKAAGVQIVLVGSM
                GGTNINHPLNSIGNANILVNRKAEDYLAGSGIPYTIIRAGGLGDKDGGIRELLVGSM
                DELLETETRTIARADVAEVCVQALQLEEEAKFKALDLASKPEGTGTPTKDFKALFTQVT
                TKF"
```

```
mRNA            complement(join(<53450. .54129,54196. .54289,54371. .54453,
                54545. .54596,54711. .54808,54958. .>56602))
                /gene="F13M22.17"
gene            complement(<53450. .>56602)
                /gene="F13M22.17"
                /note="predicted by grail"
CDS             complement(join(53450. .54129,54196. .54289,54371. .54453,
                54545. .54596,54711. .54808,54958. .56602))
                /gene="F13M22.17"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236249"
                /translation="MEDGNKLDRKKTMTMWEGLGDFEDEDDDRFFETHDRLSSALTF
                DMANASSDEDEDFDDCRLSFSSAVSSLTTASRKFRTPAMSPDYDIWAAPGSISERR
                RRLLHGMGLASNKDMVSAVSIRRVVSNAPVVSNGEEKKMKKMNHEVDEQDHDHVPVM
                LARSRSESDIERFLIEKRRKEILGKISKQRLTRTYSTICPTRTRICQYQTPIRQTPA
                VCGNGKALRGGGEALTSVMSNARVGAFFLIKNLDTGKEFIVNEYDEDGMWNRLSDLQT
                GKQLTLEEFEKCVGYSPVVKELMRRENVRINYEPLMDLRKFNSVLSKSVRLSKRRGA
                ALLKNIKGVAHSMSLRVADKVDSDGSTDSPFKKGKDHKHGKANEWVKVRPTGKSYKELS
                ALHMCOEIQAHEGAVWTIKFSQDAHYLASGGADRVHVWEVQECELMSNEGSLTPIH
                PSLCDSSGNEITVVEKKKKGKGSSGRRHNHIPDYHVPETVFVEFSFSDKPVCSLKGHLDA
                ILDLSWSKSQLLLSSSMDKTVRLWDIETKTCLKLFAHNDYVTCIQFSPVDENYFLSGS
                LDAKIRIWSIDDRHVVEWSDLHEMVTAACYTPDGQQALIGSHGICRAYDTEDCKLSQ
                TNQIDVQSNKKSQAKRKITSFQFSPVNPSEVLVTSADSRIRILDGSEVIHKFKGFRNT
                CSQLSASYSQDGKYIICASEDSQVILWKNDFHRTRSTLTTQSHEHFCKDVSAAVPWH
                GHVRGEPPPVOIHSKRHSKRISTSSOPSSTISSPTKEETSATGPTTSNRNKKSGLPPM
                PKKAATKSQIOPEEEAGPELGSSESFRSSMNSSEOHSSRFGESPSINTSSRLSSWSWF
                DSGGHGPQTIQPTAWGMVIVTATVHGEIRSYQNFGLPRRIGRQTTLF"
repeat_region   complement(56145. .56194)
                /rpt_family="(GAA)n"
misc_feature    56686. .56744
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    complement(56743. .56808)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    complement(57092. .57128)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    57429. .57497
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
misc_feature    complement(57759. .57834)
                /note="exon predicted by xgrail, quality excellent"
mRNA            complement(join(<57887. .58388,59719. .59846,60078. .60132,
                60218. .60291,60384. .60527,61543. .>61581))
                /gene="F13M22.18"
gene            complement(<57887. .>61581)
                /gene="F13M22.18"
                /note="predicted by genefinder"
CDS             complement(join(57887. .58388,59719. .59846,60078. .60132,
                60218. .60291,60384. .60527,61543. .61581))
                /gene="F13M22.18"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236250"
                /translation="MPVRVVESNTPAQAFSGTQNVSNQQKEEAWRVNVQIQGIDLEHG
                YLCGTMEALNVPMADTPVITFWEGEIVDGKNYTFYTGKWEATREDDMRHWSKFPSFSP
                LQRWKEQYFVNVGTDCGLTIAGFYVCFSCSDGSISGFYDPNSSFHHMIISSSKNVL
                KMEEVEVSKKRKFQTDQSDELSLLPLSKHTCFANVACSENTNGNSEIDTEYSMSSYVN
                STTSMECNDIEMKEESSGSCGEDKMISFESHLDYIYGTQNLEDFSEKVIENILYLDE
                QEEEEDAKGCSSNAAKFVLSSGRWTVNQGKENRILV"
repeat_region   58770. .58795
                /rpt_family="(GA)n"
repeat_region   59165. .59234
                /rpt_family="AT_rich"
repeat_region   59249. .59912
                /rpt_family="AT_rich"
repeat_region   complement(59446. .59478)
                /rpt_family="AT_rich"
misc_feature    complement(61713. .61851)
                /note="exon predicted by xgrail, quality good_shadowexon"
mRNA            join(<62177. .62406,62814. .63035,63131. .63242,63337. .63399,
                63504. .63642,63940. .63995,64072. .64145,64247. .64394,
                64470. .64652,64757. .64844,64948. .65058,65257. .65354,
```

```
gene            65591..65708,65788..65882,66002..66108,66213..>66306)
                /gene="F13M22.19"
                <62177..>66306
                /gene="F13M22.19"
CDS             join(62177..62406,62814..63035,63131..63242,63337..63399,
                63304..63642,63940..63995,64072..64145,64247..64394,
                64470..64652,64757..64844,64948..65058,65257..65354,
                65591..65708,65788..65882,66002..66108,66213..66306)
                /gene="F13M22.19"
                /codon_start=1
                /product="putative phosphoribosylaminoimidazole
                carboxylase"
                /db_xref="PID:g3236251"
                /translation="MLLLKQSSAAVLVVGNTTPVLHTSRSTYRVGPPVTRTQSFQSL
                TMANLQKLPTSSSGKLNTASAVPCSSHDASPISENRENKHVHGVSEKIVGVLGGGQLG
                RMLCQAASQLAIKVMILDPSKNCSASALSYGHMVDSFPDDSATVEEFAKRCGVLTVEIE
                HVDVDTLEKLEKQGVDCQPKASTIRIIQDKYMQKVHFSQHGIPLPEFMEISDIEGARK
                AGELFGYIPLMIKSKRLAYDGRGNAVANNQDELSSAVTALGGFSRGLYIEKWAPFVKEL
                AVIVARGRDGSMVCYPVVETIHRDNICHIVKAPADVPWKINKLATDVAQKAVGSLEGA
                GVFAVEFLFTEDSQILLNEVARPRPHNSGHQTIECCYTSQFEQHLRAVVGLPLGDPSMR
                TPASIMYNNLGEDDVIDGEAGFKLAHRLIARALCIPGASVHWTDKPEMRKQRKMGHIT
                LVGQSMGIILEQRLQCILSEQSHQVHETPRVAIIMGSDTDLPVMKDAAKILDLFGVTHE
                VKIVSAHRTPEMMYTYATSAHSRGVQVIIAGAGAAHLPGMVASLTPLPVIGVPVRAT
                RLDGVDSLLSIVQMPRGVPVATVAINNATNAALLAVRMLGISDTDLVSRMRQYQEDMR
                DENLNKGEKLETEGWESYLNQ"
mRNA            complement(join(<66647..66829,67056..67229,67307..67405,
                67776..67976,68080..68187,68273..68651,68917..69175,
                69620..70064,70183..>70242))
                /gene="F13M22.20"
gene            complement(<66647..>70242)
                /gene="F13M22.20"
CDS             complement(join(66647..66829,67056..67229,67307..67405,
                67776..67976,68080..68187,68273..68651,68917..69175,
                69620..70064,70183..70242))
                /gene="F13M22.20"
                /codon_start=1
                /product="CER1-like protein"
                /db_xref="PID:g3236252"
                /translation="MASRPGFLTDWPWTPLGSFKYLLLAPLVFDSIYSYATIRDHEKL
                LIVAVTVWRIVHSQIWISLSRYQTAKGTRKILNKSIEFDQVDRERTWDDQIIFNTLIV
                YLTKVYVSGTSTIPFWRTDGVILVALLHAGPVEFIYYWFHRALHHFLYSRYHSHHHS
                SIVTEPITLCAINSKPWVLIVAVVHPFAEHIGYTLILGLPLITTFMCGTVSVVSIALY
                LTYIDFMNNMGHCNFELIPKFLFLSLLPPLKFLCYTPSFHSLIHTQFRTNYSLFMPMYD
                YIYGTTDECSDSLYETSLEKEEEKPDAIHLTHLTSLDSIYHLRLGFASLSSHPLSSRC
                YLFLMKPFALILSFILRSFSFQFFVVERNRFRDLTLHSHLLPKFSSHYMSHQKECIN
                KMIEAAILEADKKGVKVMSLGLLNQGEELNGYGEMYVRRHPKLKIRIVDGGSLAAEVV
                LHSIPVGTKEVLFRGQITKVARAIVFSLCQNAIKVMVLRKEHSMLAEFLDDKCKENL
                VLTTNYIPMIWLVGDGLSTKEQKMAKDGTLFLPFSQFPKTLRKDCFYHTTPAMIIPH
                SAQNIDSCENWLGRRVMSAWRVGGIVHALEGWKEHECGLDDNSIINPPRVWEAALRNG
                FQPLLLPSLET"
repeat_region   complement(67680..>67715)
                /rpt_family="AT_rich"
misc_feature    complement(70565..>70634)
                /note="exon predicted by xgrail, quality marginal"
mRNA            complement(70808..>73053)
                /gene="F13M22.21"
gene            complement(70808..73053)
                /gene="F13M22.21"
                /note="highly similar to GB:X95909(polymorphism exists
                between nucleotides 73005 and 72085 which is a GA repeat.
                We found 6 copies in our sequence whereas only 5 copies
                exist in GB:X95909)"
                /gene="F13M22.21"
CDS             complement(71000..>73027)
                /gene="F13M22.21"
                /codon_start=1
                /product="receptor-like protein kinase"
                /db_xref="PID:g3236253"
                /translation="MFLKLLTIFFFFFNLIFQSSSQSLNFAYNNGFNPPTDLSIQGI
                TTVTPNGLLKLNTTVQKTGHAFYTKPIRFKDSPNGTVSSFSTSFVFAIHSQIAILSG
                HGIAFVVAPNASLPYGNPSQYIGLFNLANNGNETNHVFAVELDTILSTEFNDTNDNHV
                GIDINSLKSVQSSPAGYWDEKGQFKNLTLISRKPMQVWVDYDGRTNKIDVTMAPFNED
                KPTRPLVTAVRDLSSVLLQDMVYGFSSATGSVLSEHYILGWSFGLNEKAPPLALSRLP
                KLRFEPKRISEFYKIGMPLISLFLIFSFIFLVCYIVRRRRKFAEELEEWEKEFGKNR
                FRFKDLYYATKGFKEKGLLGTGFGSVYKGVVMPGTKLEIAVKRVSHESROGMKEFVAE
                IVSIGRMSHRNLVPLLGYCRRGELLLVYDYMPNGSLDKYLYINTPEVTLNWKQRIKVI
                LGVASGLFYLHEEWEQVVIHRDVKASNVLLDGELNGRLGDFGLARLYDHGSDPQTTHV
                VGTLGYLAPEHTRTGRATMATDVFAFGALELVACGRRPIEFQOETDETFLLVDWVFG
                LWNKGDILAAKDPNMGSECDEKEVEMVLKLGLLCSHSDPRARPSMRQVLHYLRGDAKL
                PELSPLDLSGSGMFGVHDGFSELGMSYSSSVFKGFTGGSSIADSQLSSGGR"
                72962..73005
                /rpt_family="(GAA)n"
                complement(73447..73473)
                /rpt_family="AT_rich"
repeat_region   join(<74148..74206,74511..74566,75952..>77285)
                /gene="F13M22.22"
repeat_region   <74148..>77285
                /gene="F13M22.22"
mRNA            join(74148..74206,74511..74566,75952..77285)
                /gene="F13M22.22"
gene            /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236254"
CDS             /translation="MVQIVPHVANAHMVKTSRKGLLALVVVGVEECCPTRRCDLRDPN
                LVDSNCLLSRSTHQSSLVEDSIIINHFRILNFNNRASIFLCLRIGVICSTQSKMYGG
                KSYILKGVSLSLILLLVTLLVSAKNQPRISLIEVKTCNLAKGEWVEDKKRPLYS
                GFECKQWLSNIFSCRVMGRPDFSFEGYRWQPEGCNIPEFNRVNFLRRMQNKTIAFIGD
                SLGREQFQSLMCMATGGKESPEVQNVGSEYGLVIPKGAPRPGGWAYRFPTNTTVLSY
                WSASLTDLVPMNNTDPPHLIAMHLDRPPAFIRNYLHRFVLVLNTGHNWVMHVNGTRVEGGYFKNVENAKIFTIHSLVKWLDAQLPLHPRLKAFFTTISPRHEKNH
                WVMHVNGTRVEGGYFKNVENAKIFTIHSLVKWLDAQLPLHPRLKAFFTTISPRHEKCN
                NTIPLSRGSKITGEGGSLDTIVESAVNGTRVKILDITALSKLRDEAHIAGCKLKPKKA
                SNVTSAPTFNDCLHWCLPGIPDTWNELLIAQL"
repeat_region   complement(75409..>75618)
                /rpt_family="AT_rich"
repeat_region   complement(77388..77424)
                /rpt_family="AT_rich"
repeat_region   complement(77464..77524)
                /rpt_family="AT_rich"
misc_feature    77550..77635
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
mRNA            complement(join(<78252..78416,78604..79152,79555..79945,
                80111..>80604)
                /gene="F13M22.23"
gene            complement(<78252..>80604)
                /note="predicted by genscan"
                /gene="F13M22.23"
CDS             complement(join(78252..78416,78604..79152,79555..79945,
                80111..80604))
                /gene="F13M22.23"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236255"
                /translation="MAIFDFKLHHKPSSIFPATTVKPSHVLSLVAKFFTICIFISV
                AMISFYIIFFGCSDCHHITGLRRLGPENVNLTTNSSASSTIRRPNQSSEATDISHIA
                FGIGGSIQTWRDRSRYSELWWRPNVTRGFIWLDEEPPLNMTWLSTSPPYQVSADTSRF
                SYTCWYGSRAIRMARIIKETFELGLTDVRWFIMGDDDTVFFVDNLITVLNKYDHNQM
                YYIGGNSESVEQDIVHSYAMAYGGGGIAISYPLAVELVKLLDGCIDRYASLYGSDQKI
                EACLSEIGVPLTKELGFHQVDIRGNPYGLLAAHPVAPLVTLHHLDYVPIFPGTQID
                ALRRIVSAYKTDPSRIIQHSFCHDQTRNWVSVSWGYTIQIYPTLVTAKELETPFTF
                KSWRTSSEPFSEDTRPISEDPCERPLVFFLDRVYEVGSGQTLTTYRKHVEVGESTQC
                NSPDYSRANPVEFIDVSTTLTPDLWKMAPRRQCCEIVNSEEDSESSINVKIRHFNPL
                ESVTPQSKHPNDEWTDSVLSTETI"
repeat_region   complement(79324..79361)
                /rpt_family="AT_rich"
repeat_region   80698..80727
misc_feature    complement(81151..81260)
                /note="exon predicted by xgrail, quality excellent"
misc_feature    82037..82089
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
misc_feature    82508..82558
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
mRNA            complement(<83772..>84686)
                /gene="F13M22.24"
```

```
gene            complement(<83772..>84686)
                /gene="F13M22.24"
CDS             complement(83772..84686)
                /gene="F13M22.24"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236256"
                /translation="MEKPGGFWIPKKSNKESSWEELAFAEDDAAGSLWPPRSYTCSFC
                RREFKSAQALGHMNVHRRDARLKQADDQLFPKSSSSPEYPSHKDSDNIHETSCYT
                LVFNTKPNYFKTQHSCVIDLSSSSLPYLTPSRVSSGLPGKQHTSSSPPSFVVEPSKN
                SKYIPSSSPWSCPSTVVEQKSCDLYEIPAMEGEKKRKTESDVPKIGHKAKLSLGNTTD
                LSVSMNLVIHQSFPITAHGSDEIGRVDIRKRRRRHESPSQQSIFISSLSCKSDIITR
                NEESKHKGDREFEDLDELRLGTDPPKGI"
repeat_region   84230..84321
                /rpt_family="(GAA)n"
repeat_region   84910..84962
                /rpt_family="AT_rich"
repeat_region   85196..85220
                /rpt_family="(TA)n"
repeat_region   complement(85246..85307)
                /rpt_family="AT_rich"
misc_feature    complement(85742..85810)
                /note="exon predicted by xgrail, quality marginal"
repeat_region   86102..86122
                /rpt_family="AT_rich"
misc_feature    86195..86308
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   86660..86687
                /rpt_family="AT_rich"
misc_feature    86972..87002
                /note="exon predicted by xgrail, quality marginal"
mRNA            87049..87559
                /gene="F13M22.25"
gene            87049..87569
                /gene="F13M22.25"
CDS             87120..87332
                /gene="F13M22.25"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236257"
                /translation="MQIRYGNESQQLSSSYYVRLPGDSGRFVRSDLDVSGPGSNSSNR
                ASSSTSTTTFAVYAGVQSPVSRQVVT"
mRNA            join(<88062..88169,88258..88308,88608..88684,88770..88857,
                88934..89209,89301..>89585)
                /gene="F13M22.26"
gene            <88062..>89585
                /gene="F13M22.26"
CDS             join(88062..88169,88258..88308,88608..88684,88770..88857,
                88934..89209,89301..89585)
                /gene="F13M22.26"
                /codon_start=1
                /product="putative alcohol dehydrogenase"
                /db_xref="PID:g3236258"
                /translation="MAAPIRFFELNTGAKLPCVGLGTYAMVATAIEQAIKIGYRHIDC
                ASIYGNEKEIGGVLKKLIGDGFVKREELFITSKLWSNDHLPEDVPKALEKTIQDLQID
                YVDLYLIHWPASLKKESLMPTPEMLTKPDITSTWKAMEALYDSGKARAIGVSNFSSK
                LTDLLNVARVTPAVNQVECHPVWQQGLHELCKSKGVHLSGYSPLGSQSKGEVRLKVL
                QNPIVTEVAEKLGKTTAQVALRWGLQTGHSVLPKSSSGARLKENLDVFDWSIPEDLFT
                KFSNIPQASSFIFLFFSR"
misc_feature    89723..89798
                /note="exon predicted by xgrail, quality excellent"
misc_feature    complement(89743..89842)
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
repeat_region   complement(90572..90624)
                /rpt_family="AT_rich"
mRNA            join(<90955..91074,91214..91264,91485..91561,91649..91736,
                91942..92217,92291..>92551)
                /gene="F13M22.27"
gene            <90955..>92551
                /gene="F13M22.27"
CDS             join(90955..91074,91214..91264,91485..91561,91649..91736,
                91942..92217,92291..92551)
                /gene="F13M22.27"
                /codon_start=1
                /product="putative alcohol dehydrogenase"
                /db_xref="PID:g3236259"
                /translation="MANAITFFKLNTGAKFPSVGLGTWQASPGLVGDAVAAVKIGYR
                HIDCAQIYGNEKEIGAVLKKLFEDRVVKREDLFITSKLWCTDHDPQVPEALNRTLKD
                LQLEYVDLYLIHWPARIKKGSVGIKPENLLPVDIPSTWKAMEALYDSGKARAIGVSNF
                STKKLADLLELARVPPAVNQVECHPSWRQTKLQEFCKSKGVHLSAYSPLGSPGTTWLK
                SDVLKNPILNMVAEKLGKSPAQVALRWGLQMGHSVLPKSTNEGRIKENFNVFDWSIPD
                YMFAKFAEIEQVRS"
repeat_region   complement(91371..91445)
                /rpt_family="(CAAAA)n"
misc_feature    92640..92726
                /note="exon predicted by xgrail, quality excellent"
misc_feature    complement(93519..93682)
                /note="exon predicted by xgrail, quality good_shadowexon"
misc_feature    93640..93819
                /note="exon predicted by xgrail, quality excellent"
repeat_region   complement(93688..93779)
                /rpt_family="(CAT)n"

BASE COUNT      29945 a     17475 c     16793 g     29632 t
ORIGIN

Query Match     29945 a     17475 c     16793 g     29632 t
Best Local Similarity    11.4%;      Score 70.4;   DB 9;   Length 93845;
Matches      89;  Conservative   0;  Mismatches  31;  Indels   0;   Gaps   0;

Qy     174  atctcttggcctccaagatcttatacatgtagcttttgtaaagggaattagatctgc  233
            || || |||||||| |||||||||| |||||| |||||||| ||||||| ||| |||
Db   84594  ATCTTGTGGCCGCCAAGATCTTATACGTGTAGCTTCTCGCCGGAGAGATTTAAATCGGC  84535

Qy     234  tcaagctcttggtggacacatgaagttcatagaagagtagagccatttttgagacaatc  293
            ||| |||||| || ||| |||  || ||||| |||||||  |||||| |  |||||||
Db   84534  TCAAGCCCTCGGTGGCCATATGAACGTTCACCGAAGACACAGAGCACCGGCTTAAGCAAGC  84475

RESULT    5
ATAC004684/c
LOCUS       ATAC004684    93845 bp    DNA             PLN       17-JUN-1998
DEFINITION  Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence,
            complete sequence.
ACCESSION   AC004684
NID         g3236234
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
            Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
            eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,
            Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
            Somerville,C.R. and Venter,J.C.
  TITLE     Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D. and Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
REFERENCE   3  (bases 1 to 93845)
  AUTHORS   Rounsley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jun 17, 1998 this sequence version replaced gi:3172168.
            Address all correspondence to:
            Steve Rounsley
            The Institute for Genomic Research
            9712 Medical Center Dr,
            Rockville, MD 20850,
            USA
```

```
            e-mail: rounsley@tigr.org
            BAC clone F13M22 is from Arabidopsis chromosome II and is near the
            molecular marker ve018.
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL (available by anonymous ftp
            from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
            complete sequence against a peptide database and the Arabidopsis
            EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
            genomic sequence that are not annotated as misc features.
            exons by GRAIL are annotated as misc features.
FEATURES     Location/Qualifiers
     source      1..93845
                 /organism="Arabidopsis thaliana"
                 /cultivar="Columbia"
                 /db_xref="taxon:3702"
                 /chromosome="II"
                 /map="ve018"
                 /clone="F13M22"
     misc_feature complement(10..29)
     misc_feature 168..249
     mRNA        <464..>619
                 /gene="F13M22.1"
                 /note="exon predicted by xgrail, quality marginal"
     gene        <464..>619
                 /note="exon predicted by xgrail, quality good_shadowexon"
                 /gene="F13M22.1"
     CDS         <464..619
                 /gene="F13M22.1"
                 /note="hypothetical protein, 5' partial"
                 /codon_start=1
                 /db_xref="PID:g3236260"
                 /translation="SRGDVREQRTHKDGEDKRLLDPATRRRKLETNAIIASGTKGF
                 EIGNGNL"
     mRNA        complement(join(835..1212,1373..1513,1600..1693,
                 1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
                 2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
                 4167..4353,4486..4549,4633..4842,4938..5006,5088..>5654))
                 /gene="F13M22.2"
     gene        complement(835..>5654)
                 /gene="F13M22.2"
     CDS         complement(join(1099..1212,1373..1513,1600..1693,
                 1789..1845,1933..1994,2162..2407,2492..2548,2640..2697,
                 2855..3038,3239..3405,3479..3531,3790..3871,3980..4045,
                 4167..4353,4486..4549,4633..4842,4938..5006,5088..5654))
                 /gene="F13M22.2"
                 /note="unknown protein"
                 /codon_start=1
                 /db_xref="PID:g3236235"
                 /translation="MPKEENGQLKRDRLDDDTDEGNKGDHFPSKKQAKEASNDDITSE
                 ISNPVASPVESTSLFRDVSSQPVKSGLVECSGSDFGSEETVSDDASVVGSSQTEQSSD
                 VLPSRFVLEIPKHLSSTGITKITFKLSKPKKEFDDLPLIKDHTWDAGVVKMPKKIVS
                 LSYPSNVKKLLETGILEGARVKYISTPPVRQLLGIIHSGGYLCGCTTCNFSKVLSAYE
                 FEQHAGAKTRHPNNHIFLENRRAVYNIVQELKTAPRVVLEEVIRRNVAGSALNEEGLRA
                 WKGDGGQVSNRVIACLTGIILRIIPLSVNALLSYHFVLLSETDSCESCSYIGPGLDES
                 QSLTPCSVENHYFSEKTYAKDTLDEPKRIAKKLTSHVSGTGCHKKVSEGSNRKRDNDL
                 HRLLFMPNGLPDGTELAYYVKTQISPSQFEAHAGMAARRQPYRHIFISSGSLSLHDIAM
                 SLANGHVITTGDSDDMCSICGDGGDLLLCAGCPQAFHTACLKFQSMPEGTWVCSSCND
                 GPISSKATTDPSGNARPIVIRLSRVVKAPESDIGGCVFCRSHDFSIGKFDDRTVL
                 CDQCEKEYHVGCLRENGFCDLKEIPQEKVRCCSNCSRIHTAVQSVNGCGPQLPTPLL
                 DMICRKDREKGIFTDIGDTVEWRILSGKSRYPEHLPLSRAVIFRCFDPIVAKSGR
                 DLIPVMVYGRNISGQEFGGMYCLVLIVNSLVVSAALRIFGQEVAELPIVATSREYQG
                 RGYFQGLYACVENLLSSLNVENLVLPAAEEAESIWTKKFGFTKMSDQLQEEVQKEVQL
                 TIFKGTSMLEKKVPKATGLSESTTLI"
     repeat_region 6010..6135
                 /rpt_family="(GAA)n"
     misc_feature 6023..6202
                 /note="exon predicted by xgrail, quality
                 marginal_shadowexon"
     repeat_region 6179..6236
                 /rpt_family="POLY_A"
     mRNA        join(<6647..6702,7161..7304,7322..>7499)
                 /gene="F13M22.3"
     gene        <6647..>7499
                 /note="predicted by genscan"
                 /gene="F13M22.3"
     CDS         join(6647..6702,7161..7304,7322..7499)
                 /gene="F13M22.3"
                 /note="hypothetical protein"
                 /codon_start=1
                 /db_xref="PID:g3236236"
                 /translation="MTWKKVREENKSFQFLFKFAWRIPHFVYGATWTLFLTLTVAII
                 SLAPEFAFVSAIFPSSSEVFSRRHIGSAGSSLGGLVFTGESFPEVEDGPRCAAGFCSN
                 CCGSFRRRKDYGPLGGRRSPLI"
     mRNA        complement(join(<7761..7931,8019..8065,8154..8238,
                 8323..8401,8494..8647,8824..8903,9001..9232,9769..>9886))
                 /gene="F13M22.4"
     gene        complement(<7761..>9886)
                 /gene="F13M22.4"
     CDS         complement(join(7761..7931,8019..8065,8154..8238,
                 8323..8401,8494..8647,8824..8903,9001..9232,9769..9886))
                 /gene="F13M22.4"
                 /codon_start=1
                 /product="putative ribotol dehydrogenase"
                 /db_xref="PID:g3236237"
                 /translation="MGIYGVMTGKKGKSGFGSASTAEDVTQAIDASHLTAIITGGTSG
                 IGLEAARVLAMRGAHVIAARNPKAANESKEMILQMNPNARVDYLQIDVSSIKSVRSF
                 VDQFALNVPLNILNNAGVMFCPFKLTEDGIESQFATNHIGHFLLTNLLDKMKSTA
                 RESGVQGRIVNLSSIAHTYTYSEGIKFQGINDPAGYSERRAYGQSKLSNLLHSNALSR
                 RLQEEGVNITINSVHPGLVTTNLFRYSGFSMKVFRAMTFLFWKNIPQGAATTCYVALH
                 PDLEGVTGKYFGDCNIVAPSKFATNNSLADKLWDFSVFLIDSISK"
     repeat_region 8658..8684
                 /rpt_family="AT_rich"
     repeat_region complement(9842..9915)
                 /rpt_family="(GA)n"
     misc_feature 10503..10670
                 /note="exon predicted by xgrail, quality
                 marginal_shadowexon"
     repeat_region 10725..10746
                 /rpt_family="AT_rich"
     mRNA        complement(join(11416..11526,11608..11610,12116,12551..12628,
                 12737..13604))
                 /gene="F13M22.5"
     gene        complement(11416..13604)
                 /gene="F13M22.5"
     CDS         complement(join(11610..12116,12551..12628,12737..13522))
                 /gene="F13M22.5"
                 /codon_start=1
                 /product="putative ARF1 GTPase activating protein"
                 /db_xref="PID:g3236238"
                 /translation="MAAARRLRTLQSOPENKVCVDCSQKNPQWASISYGIFMCLECSG
                 KHRGLGVHISFVRSVTMDSWSEIQIKKMDAGGNERLNNFLAQYGISKETDIISKYNSN
                 AASVYRDRIQALAEGRQWRDPPIVKESVGGLMNKKPPLSQGGGRDSGNGGWDNWDND
                 DSFRSTDMRRNQSAGDFRSSGCAGPAKSKESEDIYSRSQLEASAANKESFFAKRMAE
                 NESKPEGLPPSQGGKYVGFGSSPGPAPRSNQGSGGGDVFSVMSEGFGRLSLVAASAAN
                 VVQTGTMEFTSKVKEGGLDQTVSETVNVVASKTTEIGQRTWGIMKGVMAISQKVEEF
                 TKEEASTWNQGNKTEGNGYIQNSGIGNKTANSSFGGSQSSSGHNNSYRNSNSWDDWG
                 EENNSKKEAAPKVSTSNDDDGGWAGWDDNDAKDDDFYYQPASDKKSVGHNGKSDTAW
                 TGGGFL"
     repeat_region 11697..11757
```

Left column:

```
misc_feature    /rpt_family="(CAT)n"
                13716..13952
misc_feature    /note="exon predicted by xgrail, quality
                excellent_shadowexon"
                complement(14172..14309)
mRNA            /note="exon predicted by xgrail, quality marginal"
                join(14707..14911,15556,15765,15858..15964,16057..16189,
                16311..16602,16755..17294)
                /gene="F13M22.6"
gene            14707..17294
                /note="identical to GB:U40259"
                /gene="F13M22.6"
CDS             join(15634..15765,15858..15964,16057..16189,16311..16602,
                16755..17182)
                /gene="F13M22.6"
                /codon_start=1
                /product="origin recognition complex protein"
                /db_xref="PID:g3236239"
                /translation="MEDIENIEDEYGESRNYFLAKELGGASKRSAHKLSDIHIVDEQ
                ELRETASTIEMKHSKEISELMSDYKTMYSKWVFELRCCFGLLMYGFGSKKALVEDFAS
                ASLTDYSVVVINGYLPPSVNLKQVLLALAELLSELLKCKRKSSGSLSKCQETFPSRSMD
                DILSFLHGPGPQSGDKDCFICVVVHNIDGPALRDPESGQTLARLSSCSHIRLVASIDHVN
                APLLWDKKMVHKQFNWLWHHVPTFAPYNVEGVFFPLVLAQGSTAQTAKTAAIVLQSLT
                PNGQNVFKILAEYQLSHPDEDGMPTDDLYSASRERFFVSSQVTLNSHLTEFKDHELVK
                TKRNSDGQECLNIPLTSAIRQLLLDLNQ"
mRNA            complement(<17532..>19016)
                /gene="F13M22.7"
gene            complement(<17532..>19016)
                /gene="F13M22.7"
CDS             complement(17532..19016)
                /gene="F13M22.7"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236240"
                /translation="MENHHPSTLLSMDSSASSHEELDLEMNNRQSLLSGPPDINLPL
                SAERSPPPWNLDACDILDVGLGSQAYETENYMSVVPKVGRKCAKRVDSIWGAWFFF
                SFYFKPALNEKSKAKIVRDSNGISGFDKSDLKLDVFLVQHDMENMYMWVFKERPENAL
                GKMQLRSYMNGHSRQGDRLFPFSVEKGFVRSHRMQRKHYRGLSNPQCVHGIELVPLPN
                LTCLDEEERKRWMELTGRDLNFTIPPEASDFGSWRNLPNTDFELERPTPSLKNPTANH
                SKKLLNGSGLNLSTQPSNHSNGEATDLSPSSHKKRKDLFSNGIHEEECCLTVNPQPPV
                IEAHQNELPTWSNEFTGAMKNVYGPVTAAKTIYEDEEGYLIIISLPFVDLNSVKVSWR
                NTLTHGIIKVSCLSTSRVPFIKRHDRTFKLTDSASEHCPPGEFVREIPLSNRIPEDAN
                IEAYYDGPGSVLEILVPKLRAGPEEHEVRVCLRPNLGGNDLMIT"
misc_feature    complement(19045..19901)
                /note="exon predicted by xgrail, quality marginal"
misc_feature    complement(19455..19545)
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
mRNA            join(<20811..21507,21667..21805,22028..22434,22931..23006,
                23118..23233,23331..>23788)
                /gene="F13M22.8"
gene            <20811..>23788
                /gene="F13M22.8"
CDS             join(20811..21507,21667..21805,22028..22434,22931..23006,
                23118..23233,23331..23788)
                /gene="F13M22.8"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236261"
                /translation="MFNNTTTSFGSGPGIVVVPTPATTVPTDFPGTTITSNSTFIII
                GPPPFAPPRSIDLTPKLFIVIAFVAVPALVYALFFNGPCSSSRRNSSSSRTSSS
                SDDTPHAFVDTPPITETVTSESGGKFHKDTHSKEIGNECSVCLMVFIDSDELRQLSE
                CKHAFHVLCIETWLKDHPNCPICRTDVSVKQQTEAPNVPVNVGNVNRSGGNRRVSAT
                SRDDDWRQGLPDASTTVSVSVRDVKSDVDSQQAHPTVEDLILJIGRLLIHTRLPIRTL
                SFCSRKPSDSSSGLAPNRNLATKSTIPRFAYLVTGTKGDGKRVKRLLKAIHHPRNYYL
                LHLDLEASDEERMELAKYVRSEKKKFENVMMGLADLVTEKGPTMLASTLHGVAILLK
                KAKDWDWFINLSASDYFLMPQDDILHIFSYLPRYLNFIEHTSNIGWKENQRARPIIID
                PGFYHLKKSGVFWAKERRSLPASFKLFMGSTSVALTRPFLEFCIWGWDNLPRTLLMYY
                PFAREFFREDDLVLDKIDIELLGQTDTGLELKIPDVVKPTVSWKRLEKLMVRLLDHENF
                RAKQCK"
repeat_region   complement(20872..20971)
                /rpt_family="(TGG)n"
```

Right column:

```
repeat_region   complement(21518..21543)
                /rpt_family="AT_rich"
misc_feature    complement(24200..24291)
repeat_region   complement(25078..25159)
                /note="exon predicted by xgrail, quality good_shadowexon"
                /rpt_family="(TA)n"
mRNA            join(<25358..25408,25622..>26563)
                /gene="F13M22.9"
gene            <25358..>26563
                /gene="F13M22.9"
CDS             join(25358..25408,25622..26563)
                /gene="F13M22.9"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236241"
                /translation="MVFSSIQAYLDSSNWQQAPPSNYNHDGTGASANGHVLRPQLQP
                QQPQQQPHPNGSGGGGGGGGSIRAGSMVDRARQANVALPEAALKCPRCESTNTKFC
                YFNNYSLTQPRHFCKTCRRYWTRGGALRNVPVGGGCRRNRRTKSNSNNNNSTATSNN
                TSFSSGNASTISTILSSHYGGNQESILSQOILSPARLMNPTYNHLGDLTSNTKTDNMS
                LLNVGGLSQDLRSIHMGASGGSLMSCVDEWRSASYHQQSMGGGNLEDSSNPNPSANG
                FYSFESPRITSASISSALASQFSSVKVEDNPYKWNVNGNCSSWNDLSAFGSSR"
repeat_region   25691..25728
                /rpt_family="(CAG)n"
repeat_region   26002..26054
                /rpt_family="(CAA)n"
misc_feature    complement(27038..27122)
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   28179..28208
                /rpt_family="AT_rich"
misc_feature    complement(28259..28328)
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
repeat_region   complement(28897..28980)
                /rpt_family="(TA)n"
repeat_region   29086..29133
                /rpt_family="AT_rich"
misc_feature    complement(29945..30023)
                /note="exon predicted by xgrail, quality good"
mRNA            complement(join(<30476..30515,30659..30858,31168..>31269))
                /gene="F13M22.10"
gene            complement(<30476..>31269)
                /gene="F13M22.10"
CDS             complement(join(30476..30515,30659..30858,31168..31269))
                /gene="F13M22.10"
                /codon_start=1
                /product="putative ribosomal protein L36"
                /db_xref="PID:g3236242"
                /translation="MTTPAVKTGLFVGLNKGHVVTRRELAPRPNSRKGKTSKRTIFIR
                KLIREVAGMAPYEKRITELKVGKDKRAILKVAKRKLGTHKRAKRKREEMSSVLRKMRS
                LGGAAAEKKM"
misc_feature    complement(31377..31401)
misc_feature    complement(32006..32035)
                /note="exon predicted by genscan"
repeat_region   32044..32161
                /note="exon predicted by xgrail, quality marginal"
                /rpt_family="(GAAA)n"
repeat_region   complement(32282..32354)
                /rpt_family="(GAAAA)n"
repeat_region   32653..32718
                /rpt_family="POLY_A"
mRNA            complement(join(<32935..33310,33630..>33739))
                /gene="F13M22.11"
gene            complement(<32935..>33739)
                /gene="F13M22.11"
CDS             complement(join(32935..33310,33630..33739))
                /gene="F13M22.11"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236243"
                /translation="MEIGNVKADVNLVLNLREAERHRYFDYFIDEILWLSFLGMSET
                SLKEEEVTLGGVKETREEECSDDKLVLDICLKRKRLDEEEEEVVEEENDGFKTPTRP
                ENRIPIVRECPPAPMKRSSEMFRGRTMYCRRRLSFLPEDDVNSFITDLQWRTTTMTIK"
```

```
repeat_region    K"
                 complement(33140..33196)
                 /rpt_family="(GAA)n"
repeat_region    33357..33398
                 /rpt_family="(GAA)n"
repeat_region    33469..33489
                 /rpt_family="AT_rich"
misc_feature     complement(34248..34296)
                 /note="exon predicted by xgrail, quality
                 marginal_shadowexon"
misc_feature     35206..35258
                 /note="exon predicted by xgrail, quality good"
repeat_region    complement(35273..35341)
                 /rpt_family="(GGAA)n"
mRNA             join(35801..35886,36023..36416,36517..37130,37242..37322)
gene             35801..37322
                 /gene="F13M22.12"
CDS              join(35827..35886,36023..36416,36517..37130,37242..37307)
                 /gene="F13M22.12"
                 /codon_start=1
                 /product="actin 3 protein"
                 /db_xref="PID:g3236244"
                 /translation="MADGEDIQPLVCDNGTGMVKAGFAGDDAPRAVFPSIVGRPRHTG
                 VMVGMGQKDAYVGDEAQSKRGILTLKYPIEHGIVNNWDDMEKIWHHTFYNELRVAPEE
                 HPILLTEAPLNPKANREKMTQIMFETFNAPAMYVAIQAVLSLYASGRTTGIVLDSGDG
                 VSHTVPIYEGYALPHAILRLDLAGRDLTDALMKILTERGYSFTTTAEREIVRDIKEKL
                 CYIALDYEQELETAKTSSSVEKNYELPDGQVITIGSERFRCPEVLYQPSMIGMENAGI
                 HETTYNSIMKCDVDIRKDLYGNIVLSGGTTMFPGIADRMSKEITALAPSSMKIKVVAP
                 PERKYSVWIGGSILASLSTFQQMWIAKAEYDESGPSIVHRKCF"
repeat_region    37700..37817
                 /rpt_family="(CAAAA)n"
mRNA             complement(<37864..>38967)
gene             complement(<37864..>38967)
                 /gene="F13M22.13"
CDS              complement(37864..38967)
                 /gene="F13M22.13"
                 /codon_start=1
                 /product="myb-related protein"
                 /db_xref="PID:g3236245"
                 /translation="MERQRWSGEEDALLRAYVRQFGPREWHLVSERMNKPLNRDAKS
                 CLERRKNYLKPGIKKGSLTEEEQRLVIRLQEKHGNKWKKIAAEVPGRTAKRLGKWWEV
                 FKEKQREEKRESNKRVEPIDESKYDRILESFAEKLVKERSNVVPAAAAATVVMANSN
                 GGFLHSEQQVOQPNPVVIPPWLATSNNGNNVVARPPSVTLTLSPSTVAAAAPQPPIPWL
                 QQQPERAENGPGGLVLGSMMPSCSGSSESVFLSELVECCRELEEGHRAWADHKKEAA
                 EAKDQKLADQWTSRHIRLTKFLEQQMGCRLDRP"
repeat_region    complement(39531..39561)
                 /rpt_family="AT_rich"
repeat_region    complement(40737..40818)
                 /rpt_family="(TAA)n"
repeat_region    complement(41088..41112)
                 /rpt_family="POLY_A"
repeat_region    complement(41190..41213)
                 /rpt_family="POLY_A"
misc_feature     complement(41283..41310)
                 /note="exon predicted by xgrail, quality
                 marginal_shadowexon"
repeat_region    complement(41558..41602)
                 /rpt_family="AT_rich"
repeat_region    complement(41966..42068)
                 /rpt_family="(GAAAA)n"
repeat_region    complement(42281..42325)
                 /rpt_family="AT_rich"
misc_feature     42889..42968
                 /note="exon predicted by xgrail, quality excellent"
misc_feature     complement(43132..43154)
                 /note="exon predicted by xgrail, quality good"
repeat_region    43371..43379
                 /rpt_family="(TAAAA)n"
mRNA             complement(join(43750..44452,45240..45558,45719..45963))

gene             complement(43750..45963)
                 /gene="F13M22.14"
CDS              complement(join(44143..44452,45240..45558,45719..45878))
                 /gene="F13M22.14"
                 /codon_start=1
                 /product="putative expansin protein"
                 /db_xref="PID:g3236246"
                 /translation="MRATAFRVGLWLAVTASFLLTATNAKIPGVYSGGPWQNAHATFY
                 GGSDASGTMGGACGYGNLYSOGYGVNTAALSTALFNGFSCGACFEIKCTDDPRWCVP
                 GNPSILVTATNFCPPNFAQPSDDGGWCNPPREHFDLAMPMFLKIGLYRAGIVPVSVRR
                 NAVLIGQSLSFFVTASDRRSSTSWNVAPATWQFGQTSGKNFRV"
repeat_region    44795..44903
                 /rpt_family="AT_rich"
repeat_region    46287..46312
                 /rpt_family="AT_rich"
misc_feature     complement(47301..47444)
                 /note="exon predicted by xgrail, quality
                 excellent_shadowexon"
repeat_region    48559..48589
                 /rpt_family="AT_rich"
mRNA             <48689..>50845
gene             <48689..>50845
                 /gene="F13M22.15"
CDS              48689..50845
                 /gene="F13M22.15"
                 /codon_start=1
                 /product="SCARECROW-like protein"
                 /db_xref="PID:g3236247"
                 /translation="MITEPSLTGISGMVNRNRLSGLPDQPSSHSFTPVTLYDGFNYNL
                 SSDHINTVVAAPENSVFIREEEEDPADDFDFSDAVLGYISQMLNEEDMDDKVCMLQ
                 ESLDLEAAERSLYEATGKYPPSPERNLAFAERNSENLDRVVPGNYTGGDCIGFGNGG
                 IKPLSSGFTLDFRNPQSCSSILSVPQSNGLITIYGDGIDESSKNNRENHQSVWLFRRE
                 IEEANRFNPEENELIVNFREENCVSKARKNSSRDEICVEERSSKLPAVFGEDILRSD
                 VVDKILVHVPGGESMKEFNALRDVLKKGVEKKASDAQGGKRRARGRGRGRGRGGGG
                 QNGKKEVVDLRSLLIHCAQAVAADDRRCAGQLLKQIRLHSTPFGDGNRLAHCFANGL
                 EARLAGTGSQIYKGIVSKPRSAAAVLKAHQLFLACCPFRKLSYFITNKTIRDLVGNSQ
                 RVHVIDFGILYGFQWPLIHRFSMYGSPKVRITGIEFPQPGFRPAQRVEETGQRLAAY
                 AKLFGVPFEYKAIAKKWDAIQLEDLDIDRDEITVVNCLYRAENLHDESVKVESCRDTV
                 LNLIGKINPDLFVFGIVNGAYNAPFFVTRFREALFHFSSIFDMLETIVPREDEERMFL
                 EMEVFGREALNVIACEGWERVERPETYKQWIVRAMSGLVQVPFDPSIMKTSLHKVHT
                 FYHKDFVIDQDNRWLLQGWKGRTVMALSVWKPESKA"
repeat_region    49591..49709
                 /rpt_family="(GAA)n"
repeat_region    49635..49689
                 /rpt_family="(GGA)n"
misc_feature     51112..51155
                 /note="exon predicted by xgrail, quality good"
misc_feature     complement(51301..51387)
                 /note="exon predicted by xgrail, quality
                 marginal_shadowexon"
mRNA             complement(join(51428..51666,51782..51905,51988..52049,
                 52131..52264,52306..52342,52427..52698,52779..>53043))
gene             complement(51428..>53043)
                 /gene="F13M22.16"
CDS              complement(join(51547..51666,51782..51905,51988..52049,
                 52131..52264,52306..52342,52427..52698,52779..53043))
                 /gene="F13M22.16"
                 /note="unknown protein"
                 /codon_start=1
                 /db_xref="PID:g3236248"
                 /translation="MAMMTTTTFHPLLPANTYKSGAVASSFVSVPRSSSLQFRSL
                 VSDSTSICGPSKFTGKNRRVSVTVSAAATTEPLTVLVTGAGGRTGQIVKLKERSEQ
                 FVARGLVRTRESKEKINGEDEVFIGDIRDTASIAPAVEGIDALVILTSAVPQMKPGFD
                 PSKGGRPEFFDDGAYPEQVDWIGQKNQIDAGVDDGLSFVVTAKAAGVKQIVLVGSM
                 GGTNIHPLNSIGNANILVWKRKAEQYLADSGIPYTIIRAGGLQDKDGIRELLVGKD
                 DELLETERTIARADVAEVCVQALQEEAKFKALDLASKPGTGTPTKDKALFTQVT
                 TKF"
mRNA             complement(join(<53450..54129,54196..54289,54371..54453,
                 54545..54596,54711..54868,54958..>56602))
```

```
gene          /gene="F13M22.17"
              complement(<53450..>56602)
              /note="predicted by grail"
CDS           /gene="F13M22.17"
              complement(join(53450..54129,54196..54289,54371..54453,
              54545..54596,54711..54868,54958..56602))
              /gene="F13M22.17"
              /note="hypothetical protein"
              /codon_start=1
              /db_xref="PID:g3236249"
              /translation="MEDGNKLDRKKTMTMNWEGLGDFEDEDDDRFFETHDRLSSALTF
              DMANASSDEDEDFDDCRLSFSSAVSSLTTASRKFRTPAMSPDYDIWMAAPGSISERR
              RRLLHGMGLASNKDMVSAVSIRRVVSNAPVVSNGEEKKMKKKMNHEVDEQDHDHPVM
              LARSRSESDIERFLIEKRRKEEIIGKISKQRLRTRYSTICPTRTRICQYQTPIRQTPA
              VCGNGKALRGGGEALTSVMSNARVGAFFLIKNLDTGKEFIVNEYDEDGMWNRLSDLQT
              GKQLTLEEFEKCVGYSPVVKELMRRENVNRINYEPLMDLRKFNSYLSKSVRLSKRRGA
              ALLKNIKGVAHSMSLRVADKDVSDGSTDSPKKGDLHKDKANEVKVRPTGKSYKELS
              ALHMCQEIQAHEGAVWTIKFSQDAHYLASGGADRVIHVWEVQECELMSMNEGSLTPIH
              PSLCDSSGNEITVVEKKKKGKGSSGRRHNHIPDYVHVPETVFSFSDKFVCSLKGHLDA
              ILDLSWSKSQLLLSSMDKTVRLWDIETKTCLKLFAHNDYVTCIQFSPVDENYFLSGS
              LDAKIRIWSIQDRHVVEWSDLHEMVTAACYTPDGQGALIGSHKGICRAYDTEDCKLSQ
              TNQIDVQSNKKSQAKRKTISFQFSPVNPSEVLVTSADSRIRILDGSEVIHKFKGFRNT
              CSQLSASYSQDGKYIICASEDSQVYLWKNDFHRTRSTLTTQSHEHFHCKDVSAAVPWH
              GHVRGEPPVQIHSKRHSKRISTSSQPSTISSPTKEETSATGPTTSNRNKKSGLPPM
              PKKAATKSQIQPEEEAGPELGSSESFRSSMNSSEQHSSRFGESPSINTSSRLSSWSWF
              DSGGHGPQTIQPTAWGMVIVTATVHGEIRSYQNFGLPRRIGRQTLF"
repeat_region complement(56145..56194)
              /rpt_family="(GAA)n"
misc_feature  56686..56744
              /note="exon predicted by xgrail, quality
              marginal_shadowexon"
misc_feature  complement(56743..56808)
              /note="exon predicted by xgrail, quality marginal"
misc_feature  complement(57092..57128)
              /note="exon predicted by xgrail, quality marginal"
misc_feature  57429..57497
              /note="exon predicted by genefinder"
misc_feature  complement(57759..57834)
              /note="exon predicted by xgrail, quality
              marginal_shadowexon"
mRNA          complement(join(<57887..58388,59719..59846,60078..60132,
              60218..60291,60384..60527,61543..>61581))
              /gene="F13M22.18"
gene          complement(<57887..>61581)
              /gene="F13M22.18"
              /note="predicted by genefinder"
CDS           complement(join(57887..58388,59719..59846,60078..60132,
              60218..60291,60384..60527,61543..61581))
              /gene="F13M22.18"
              /note="hypothetical protein"
              /codon_start=1
              /db_xref="PID:g3236250"
              /translation="MPVRVVESNTPAQAFSGTQNVSNQQKEEAWRVNVQIQGIDLEHG
              YLCGTMEALNVPMADTPVITFWEGEIVDGKNYTFYTGKWEATREDDMRHWSKPFSFSP
              LQRWKEQYFVNVGTDCGLITIAGFYYVCFSCSDGSISGFYYDPNSSFHHMIISSSKNVL
              KMEEVSVKKRKFQTDQSDELSLLPLSKHTCFANVACSENTNGNSEIDTEYSMSSYVN
              STTSMECNDIEMKEESSGSCGEDKMISFESHLDYIYGTQNLEDFSEKVIENILYLDE
              QEEEEDAKGCSSNAAKFVLSSGRWTVNQGKENRILV"
repeat_region 58770..58795
              /rpt_family="(GA)n"
repeat_region 59165..59234
              /rpt_family="AT_rich"
repeat_region 59249..59312
              /rpt_family="AT_rich"
repeat_region complement(59446..59478)
              /rpt_family="AT_rich"
misc_feature  complement(61713..61851)
              /note="exon predicted by xgrail, quality good_shadowexon"
mRNA          join(<62177..62406,62814..63035,63131..63242,63337..63399,
              63504..63642,63940..63995,64072..64145,64247..64394,
              64470..64652,64757..64844,64948..65058,65257..65354,
              65591..65708,65788..65882,66002..66108,66213..>66306)
              /gene="F13M22.19"


gene          <62177..>66306
              /gene="F13M22.19"
CDS           join(62177..62406,62814..63035,63131..63242,63337..63399,
              63504..63642,63995..64072..64145,64247..64394,
              64470..64652,64757..64844,64948..65058,65257..65354,
              65591..65708,65788..65882,66002..66108,66213..66306)
              /gene="F13M22.19"
              /codon_start=1
              /product="putative phosphoribosylaminoimidazole
              carboxylase"
              /db_xref="PID:g3236251"
              /translation="MLLLKQSSAAVLVVGNTTVLHTSRSTYRVGPFVPVTRTQSFQSL
              TMANLQKLPTSSSGKLNTASAVPCSSHDASPISENRENKHVHGVSEKIVGVLGGGQLG
              RMLCQAASQLAIKYMILDPSKNCSASALSYGHMVDSFDDSATVEFAKRCGVLTVEIE
              HVDVDTLEKLEKQGVDCQPKASTIRIIQDKYMQKVHFSQHGIPLPEFMEISDICEGAK
              AGELFGYPLMIKSKRLAYDGRGNAVANNQDELSSAVTALGGFSRGLYIEKWAPFVKEL
              AVIVARGRGSMVCYPVVTHRDNICHIVKAPADVPWINKLATDVAQKAVGSLEGA
              GVFAVELFTEDSQILLNEVAPRPHNSGHQTIECCYTSQFEQHLRAVVGLPLGDPSMR
              TPASIMYNLGEDDVIDGEAGFKLAHRLIARALCIPGASVHWYDKPEMRKQRMGHIT
              LVGSMGILEQRLQCILSEQSHQVHETPRVAIIMGSDTDLPVMKDAAKILDLFGVTHE
              VKIVSAHRTPEMMYTATSAHSRGVQVIIAGAGGAAHLPGMVASLTPLPVIGVPVRAT
              RLDGVDSLLSIVQMPRGVPVATVAINNATNAALLAVRMLGISDTDLVSRMRQYQEDMR
              DENLNKGEKLETEGWESYLNQ"
mRNA          complement(join(<66647..66829,67056..67229,67307..67405,
              67776..67976,68080..68187,68273..68651,68917..69175,
              69620..70064,70183..>70242))
              /gene="F13M22.20"
gene          complement(<66647..>70242)
              /gene="F13M22.20"
CDS           complement(join(66647..66829,67056..67229,67307..67405,
              67776..67976,68080..68187,68273..68651,68917..69175,
              69620..70064,70183..70242))
              /gene="F13M22.20"
              /codon_start=1
              /product="CERI-like protein"
              /db_xref="PID:g3236252"
              /translation="MASRPGFLTDWPWTPLGSFKYLLLAPLVPDSIYSYATIRDHEKL
              LIVAVTVWRIVHSQIWISLSRYQTAKGTRILNKSIEFDQVDRERTWDDQIIFNTLIV
              YLIKVYVSGTSTIPFWRDGVILVALLHAGPVEFIYWFHRALHHFLYSRYHSHHHS
              SIVTEPITLCATNSKPWVLIVAVVHPFAEHIGTLILGLPLITTFMCGTVSVSIALY
              LTYIDFMNMGHCNFELIPKFLFSLLPPLKFLCYTPSFHSLHHTQFRTNYSLFMNAYD
              YIYGTTDECSDSLYETSLEKEEEKPDAIHLTHLTSLDSIYHLRLGFASLSSHPLSSRC
              YLFLMKPFALILSFILRSFSFQTFVVERNRFRDLTLHSHLLPKFSSHYMSHQQKECIN
              KMIEAAILEADKKGVKRHPKLKIWSLGLLNQGEENGYGEMYVRRHPKLKIRIVDGGSLAAEVV
              LHSIPVGTKEVLPRGQITKVARAIVFSLCQNAIKVMVLRKEEHSMLAEFLDDKCKENL
              VLTNYYPMIWLVGDGLSTKEQKMAKDGTLFLPFSQFPPKTLRKDCFYHTTPAMIIPH
              SAQNIDSCENWLGRRVMSAWRVGGIVHALEGWKEHECGLDDNSIINPPRVWEAALRNG
              FQPLLPSLET"
repeat_region complement(67680..67715)
              /rpt_family="AT_rich"
misc_feature  complement(70565..70634)
              /note="exon predicted by xgrail, quality marginal"
mRNA          complement(70808..73053)
              /gene="F13M22.21"
gene          complement(70808..73053)
              /note="highly similar to GB:X95909(polymorphism exists
              between nucleotides 73005 and 72085 which is a GA repeat.
              We found 6 copies in our sequence whereas only 5 copies
              exist in GB:X95909)"
              /gene="F13M22.21"
CDS           complement(71000..73027)
              /gene="F13M22.21"
              /codon_start=1
              /product="receptor-like protein kinase"
              /db_xref="PID:g3236253"
              /translation="MFLKLLTIFFFFFNLIFQSSSQSLNFAYNNGFNPPTDLSIQGI
              TTVIPNGLLKTNTTVQKTGHAFYTKPIRFKDSPNGTVSSFSTSFVFAIHSQIALLSG
              HGIAFVVAPNASLPYGNPSQYIGLFNLANNGNETNHVFAVELDTILSTEFNDTNDNHV
              GIDINSLKSVQSSPAGYWDEKGQFKNLTLISRKPMQVWDYDGRTNKIDVTMAPFNED
              KPTRPLVTAVRDLSSVLLQDMYVGFSSATGSVLSEHYIGWSFGLNEKAPPLALSRLP
              KLPRREPKRISEFYKIGMPLISLFLISFIFLVCVIVRRRKFAEELEEWEKEFGKNR
              FRFKDLYATKGFKEKGLLGTGFGSVYKGVMPGTKLEIAVKRVSHESRGGMKEFVAE
              IVSIGRMSHRNLVPLLGYCRRGELLLVDYMPNGSLDKYLYNTPEVTLMKNQRIKVI"
```

```
repeat_region   LGVASGLFYLHEWEQVVIHRDVKASNVLLDGELNGRLGDFGLARLYDHGSDPQTTHV
                VGTLGYLAPEHTRTGRATMATDVFAFGAFLLEVACGRRPIEFQQETDETFLVDWVFG
                LWNKGDILAAKDPNMGSECDEKEVEMVLKLGLLCSHSDPRARPSMRQVLHYLRGDAKL
                PELSPLDLSGSGMFGVHDGFSELGMSYSSSVFKGFTGGSSIADSQLSGGR"
                72962..73005
                /rpt_family="(GAA)n"
repeat_region   complement(73347..73473)
                /rpt_family="AT_rich"
mRNA            join(<74148..74206,74511..74566,75952..>77285)
                /gene="F13M22.22"
gene            <74148..>77285
                /gene="F13M22.22"
CDS             join(74148..74206,74511..74566,75952..77285)
                /gene="F13M22.22"
                /codon_start=1
                /db_xref="PID:g3236254"
                /translation="MVQIVPHVANAHMVKTSRKGLLALVVVGVEECCPTRRCDLRDPN
                LVVDSNCLLSRSTHQSSLVEDSIIINHPRILNFRASIFLCLRIGVCTSQSKMYGD
                KSYILKGSVSLSLILILLVTTLLVSAKNQPRISLIEVKTCNLAKGEWVEDKKRPLYS
                GFECKQWLSNIFSCRVMGRPDFSFEGYRWQPEGCNIPEFNRVNFLRRMQNKIIAFIGD
                SLGREQFQSLMCMATGGKESPEVONVGSEYGLVIPKGAPRPGGWAYRFPTNTTVLSY
                WSASLTDLVPMNNTDPPHLIAMHLDRPPAFIRNYLHRFHVLVLNTGHHWSRDKIEKNH
                WVMHVNGTRVEGGYFKNVENAKIFTIHSLVKWLDAQLPHPRLKAFFTISPRHEKCN
                NTIPLSRGSKITGGGSLDTIVESAVNGTRVKILDITALSKLRDEAHIAGCKLKPKKA
                SNVTSAPTFNDCLHWCLPGIPDTWNELLIAQL"
repeat_region   complement(75409..75618)
                /rpt_family="AT_rich"
repeat_region   complement(77388..77424)
                /rpt_family="AT_rich"
repeat_region   complement(77464..77524)
                /rpt_family="AT_rich"
misc_feature    77550..77635
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
mRNA            complement(join(<78252..78416,78604..79152,79555..79945,
                80111..>80604))
                /gene="F13M22.23"
gene            complement(<78252..>80604)
                /gene="F13M22.23"
                /note="predicted by genscan"
CDS             complement(join(78252..78416,78604..79152,79555..79945,
                80111..80604))
                /gene="F13M22.23"
                /note="hypothetical protein"
                /codon_start=1
                /db_xref="PID:g3236255"
                /translation="MAIFDPFKLHHKPPSIFPATTVKPSHVLSLVAKFFTICIFISV
                AMISFYIIFFGCSDCHHITGLRRLGPENVLNTNSSASSTIRRPNQSSEATDISHIA
                FGIGGSIQTWRDRSRYSELMWRPNVTRGFIWLDEEPPLNMTWLSTSPPYQVSADTSRF
                SYTCWYGSRSAIRMARIIKETFELGLTDVRWFIMGDDDTVFFVDNLITVLNKYDHNQM
                YYIGGNSEVEQDIVHSYAMAYGGGGIAISYPLAVELVKLLDGCIDRYASLYGSDQKI
                EACLSEIGVTKLPKELGFHQVDIRGNPYGLLAAHPVAPLVTLHHDYVDPIFPGTTQID
                ALRRLVSAYKTDPSRIIOHSFCHDQTRNWYVSVSWGYTIQIYPLVTAKELETPFLTF
                KSWRTSSEPFSEDTRPISEDPCERPLYVFLDRVYEVGSGQILTTYRKHVEVGESTQC
                NSPDYSRANPVEFIDVTTLTPDLWKMAPRRQCCEIVNSEEDSESVINVKIRRFNPL
                ESVTPQSKHPNDEWTDSVLSTETI"
repeat_region   complement(79324..79361)
                /rpt_family="AT_rich"
repeat_region   80698..80727
                /rpt_family="AT_rich"
misc_feature    complement(81151..81260)
                /note="exon predicted by xgrail, quality excellent"
misc_feature    82037..82089
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
misc_feature    82508..82558
                /note="exon predicted by xgrail, quality
                marginal_shadowexon"
mRNA            complement(<83772..>84686)
                /gene="F13M22.24"
gene            complement(<83772..>84686)
                /gene="F13M22.24"

CDS             complement(83772..84686)
                /gene="F13M22.24"
                /codon_start=1
                /product="putative zinc finger protein"
                /db_xref="PID:g3236256"
                /translation="MEKPGGFWIPKKSNKESSWEELAFAEDDAAGSLWPPRSYTCSFC
                RREFKSAQALGGHMNVHRRDRARLKQADDQYLFPKSSSSPEYPSHKDSDNIHETSCYT
                LVFNTKPNYFKTQHSCVIDLSSSSLPYLTPSRVSSGLPGKQHTSSSPSFVVEPSKN
                SKYIPSSSPWSCPSTVVEQKSCDLYEIPAMEGEKKRKTESDVPKIGHKAKLSLGNTTD
                LSVSMNLVIHQSFPTAHGSDEEIGRVDIRKRRRHESPSQQSIFISSLSCKSDIITR
                NEESKHKGDRFEDLDLELRLGTDPPKGI"
repeat_region   84230..84321
                /rpt_family="(GAA)n"
repeat_region   84910..84962
                /rpt_family="AT_rich"
repeat_region   85196..85220
                /rpt_family="(TA)n"
repeat_region   complement(85246..85307)
                /rpt_family="AT_rich"
misc_feature    complement(85742..85810)
                /note="exon predicted by xgrail, quality marginal"
repeat_region   86102..86122
                /rpt_family="AT_rich"
misc_feature    86195..86308
                /note="exon predicted by xgrail, quality good_shadowexon"
repeat_region   86660..86687
                /rpt_family="AT_rich"
misc_feature    86972..87002
                /note="exon predicted by xgrail, quality marginal"
mRNA            87049..87569
                /gene="F13M22.25"
gene            87049..87569
                /gene="F13M22.25"
CDS             87120..87332
                /gene="F13M22.25"
                /note="unknown protein"
                /codon_start=1
                /db_xref="PID:g3236257"
                /translation="MQIRYGNESQQLSSSYYVRLPGDSGRFVRSDLDVSGPGSNSSNR
                ASSSTSTTTFAVYAGVQSPVSRQVVT"
mRNA            join(<88062..88169,88258..88308,88608..88684,88770..88857,
                88934..89209,89301..>89585)
                /gene="F13M22.26"
gene            <88062..>89585
                /gene="F13M22.26"
CDS             join(88062..88169,88258..88308,88608..88684,88770..88857,
                88934..89209,89301..89585)
                /gene="F13M22.26"
                /codon_start=1
                /product="putative alcohol dehydrogenase"
                /db_xref="PID:g3236258"
                /translation="MAAPIRFFELNTGAKLPCVGLGTYAMVATAIEQAIKIGYRHIDC
                ASIYGNEKEIGGVLKKLLIGDGFVKRELFITSKLWSNDHLPEDVPKALEKTLQDLQID
                YVDLYLIHWPASLKKESLMPTPEMLTKPDITSTWKAMEALYDSGKARAIGVSNFSSKK
                LTDLLNVARVTPAVNQVECHPVWQQGLHELCKSKGVHLSGYSPLGSQSKGEVRLKVL
                QNPIVTEVAELGKTAQVALRWGLQTGHSVLPKSSSGARLKENLDFVDWSIPEDLFT
                KFSNIPQASSFIFLFFSR"
misc_feature    89723..89798
                /note="exon predicted by xgrail, quality excellent"
misc_feature    complement(89743..89842)
                /note="exon predicted by xgrail, quality
                excellent_shadowexon"
repeat_region   complement(90572..90624)
                /rpt_family="AT_rich"
mRNA            join(<90955..91074,91214..91264,91485..91561,91649..91735,
                91942..92217,92291..>92551)
                /gene="F13M22.27"
gene            <90955..>92551
                /gene="F13M22.27"
CDS             join(90955..91074,91214..91264,91485..91561,91649..91736,
                91942..92217,92291..92551)
                /gene="F13M22.27"
                /codon_start=1
```

```
                    /product="putative alcohol dehydrogenase"
                    /db_xref="PID:g3236259"
                    /translation="MANAITFFKLNTGAKFPSVGLGTWQASPGLVGDAVAAAVKIGYR
                    HIDCAQIYGNEKEIGAVLKKLFEDRVVKREDLFITSKLWCTDHDPQDVPEALNRTLKD
                    LQEYVDYLYLIHWPARIKKGSVGIKPENLLPVDIPSTWKAMEALYDSGKARAIGVSNF
                    STKKLADDLELARVPPAVNQVECHPSWRQTKLQEFCKSKGVHLSAYSPLGSPGTTWLK
                    SDVLKNPILNMVAEKLGKSPAQVALRWGLQMGHSVLPKSTNEGRIKENFNVFDWSIPD
                    YMFAKFAEIEQVRS"
    repeat_region   complement(91371..91445)
                    /rpt_family="(CAAAA)n"
    misc_feature    92640..92726
                    /note="exon predicted by xgrail, quality excellent"
    misc_feature    complement(93519..93682)
                    /note="exon predicted by xgrail, quality good_shadowexon"
    misc_feature    93640..93819
                    /note="exon predicted by xgrail, quality excellent"
    repeat_region   complement(93688..93779)
                    /rpt_family="(CAT)n"
BASE COUNT     29945 a   16793 g   29632 t
ORIGIN

Query Match          11.4%;       Score 70.4;   DB 39;    Length 93845;
Best Local Similarity   74.2%;    Pred. No. 4e-06;   Gaps      0;
Matches     89;   Conservative    0;   Mismatches    31;   Indels   0;   Gaps    0;

Qy    174  atctcttggccctccaagatcttatacatgtagcttttgtaaaagggaattagatctgc  233
           || |||| | |||||||||  ||| ||| |||||||||  | ||| ||| ||||| ||
Db   84594  ATCTTTGTGGCCGCCCAAGATCTTATACGTGTAGCTTCTGCCGGAGAGATTTAAATCGGC  84535

Qy    234  tcaagctcttgtggtggacacatgaatgttcatgaaggagatagagccattttgagacaatc  293
           |||||| ||  || ||| || || ||| |||| ||| ||  || || || || |||| ||
Db   84534  TCAAGCCCTCGGTGGCCATATGAACGTTCACCGAAGACAGCACGCGCCTTAAGCAAGC  84475

RESULT   6
DDU32174
LOCUS       DDU32174     5931 bp    DNA            INV       22-JAN-1997
DEFINITION  Dictyostelium discoideum spore lysis A (splA) gene, complete cds.
ACCESSION   U32174
NID         g974333
KEYWORDS    .
SOURCE      slime mold.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; mitochondrial eukaryotes; Dictyostelida;
            Dictyostelium.
REFERENCE   1  (sites)
  AUTHORS   Tan,J.L. and Spudich,J.A.
  TITLE     Developmentally regulated protein-tyrosine kinase genes in
            Dictyostelium discoideum
  JOURNAL   Mol. Cell. Biol. 10 (7), 3578-3583 (1990)
  MEDLINE   90287147
REFERENCE   2  (bases 1 to 5931)
  AUTHORS   Nuckolls,G.H., Osherov,N., Loomis,W.F. and Spudich,J.A.
  TITLE     The Dictyostelium dual-specificity kinase splA is essential for
            spore differentiation
  JOURNAL   Development 122 (10), 3295-3305 (1996)
  MEDLINE   97053827
REFERENCE   3  (bases 1 to 5931)
  AUTHORS   Nuckolls,G.H. and Spudich,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-1995) Glen H. Nuckolls, Departments of
            Biochemistry and Developmental Biology, Stanford University, B400
            Beckman Center, Stanford, CA 94305, USA
FEATURES             Location/Qualifiers
     source          1..5931
                     /organism="Dictyostelium discoideum"
                     /strain="JH10"
                     /db_xref="taxon:44689"
                     /chromosome="4"
     gene            861..5763
                     /gene="splA"
     CDS             join(861..1001,1083..1155,1224..5763)
```

```
                    /gene="splA"
                    /standard_name="spore lysis A"
                    /note="exhibits tyrosine and serine/threonine
                    autophosphorylation activity; also called DPYK1"
                    /codon_start=1
                    /product="non-receptor tyrosine kinase"
                    /db_xref="PID:g974334"
                    /translation="MSFFFKFKKLVPSNEYRWDLRKSNSLTLNIEDKSRCSYRLPTSG
                    SKGIAKSTQPFSSSSFYFELFITNGNGDKICFGLTTNDHPIEVYPGNYQGSVGYSGDG
                    KCYFGTNEGRVYGPSFSSGDVVGCGYDSSSKTLYFTKNGVILGAVAQKVNLIGLYPTV
                    GLQNPGESVINFFGPFSYRGAPEKPSKQSTIKDSGGSSIIPSEDLIPKEEFVCRWS
                    EKKNYHGKHVVRNRTAFLPLDSPKDTIGGVRATQPFGEGFCYFEVIIDQLDKGQLSI
                    GLANLEYPTFYHVGWMPRSYGYHNDDGRKFRWREEPGVNEGSYGSSYKKGDIIGCGL
                    SFTSREIFFTKNGMYLGTAFSNVYGVFYPSVAFNEPGISITGVPGGFPFKFSQVTLMLK
                    NVNSTSLVPNGNNNNNSNNNNNNNNNNNNIIGNGKITTTTTSTSPSSINNNEDISSNN
                    NNNNNNNNNNNNNNNNNNNNNTTNNNINNTNNNNSNSNNNNNNN
                    SNSNSNNNNINNNNNNNNNNNNIYLTKKPSIGSTDESSTGSLGGNNSSGNNNSSSG
                    SIGNNSSIIKQRSPPHSINGPLMLPPSSTNNNNNIYSSYNSTTAGSSTTILPTLNHPI
                    FGNTTSNNNSSSTLSVGGNNNLLGRHCQSLPITASTNHTLSSLGVVSFSSPSSSPKTS
                    PRKIVNSSEDLGFVQTFQDODGQPPSAWRRCGKSIKTKDDITLTIIKKKTSVAMADRP
                    FSSNSSSTICYFEVYLEGHDKKGSIVGLSHSTYPFIKHIGREPKSYGFSSEGEKYGG
                    SEIGEPYGPFFFDGDSIASSCVIGCGINTSTRDIFFTKNGHYLGVAFSRVTSDPLYP
                    SISFRGVVGGLCVATFPGGHFRFNIEDLPGISPSVWTEALGPDRQGSGFKNWAPNDVA
                    IWLESFNYGQYRKNFRDNNISGRHLEGITHAMLKNDLGIEPYGHREDIINRLNRMIQI
                    WNDKSPDSYPKIAIDSSDKIRWPASGGSSGGINISGGVVIGSSSGSDDGITEISSSSK
                    NIRPYKSYTQKEIEDRNRRSTISGGEKKNKYIDNQMDPHQIGSMDSDGLLPDFGQGF
                    PDEKNSSKTLSNEQIRYLQQRKDEPPIAISSTGNGSVSSTGGSGFLTFPSSNSLTH
                    PPQRDKFTQEFTHLPPITSNYKGITNTGQPHKSFDQPLELFPRHSAFSNNGNNGNNN
                    NNNNNIKANQQQOQSYOQSTOQQQQHITSTSTTNKWIDFFGGWETOSSLSHP
                    PSRPPPPPPPQLPVRSEYEIDFNELEFCGTIGKFFGEVKRGYWRETDVAIKIIYR
                    DQFKTKSSLVMFQNEVGILSKLRHPNVVQFLGACTAGGEDHHCIVTEWMGGSLRQFL
                    TDHFNLLEQNPHIRLKLALDIAKGMNYLHGWTPPILHRDLSSRNILLDHNIDPKNPVV
                    SSRQDIKCKISDFGLSRLKKEQASQMTQSVGCIPYMAPEVFKGDSNSEKSDVYSYGMV
                    LFELLISDEPQQDMKPMKMAHLAAYESYRPPIPLTTSSKWKEILTQCWDSNPDSRPTF
                    KQIIVHLKEMEDQGVSSFASVPVQTIDTGVYA"
     intron          1002..1082
                     /gene="splA"
                     /note="intron confirmed by sequencing RT-PCR product"
     intron          1157..1223
                     /gene="splA"
                     /note="intron confirmed by sequencing RT-PCR product"
     conflict        5259
                     /gene="splA"
                     /note="reported as C in mRNA sequence, GenBank Accession
                     Number M33785"
                     /citation=[1]
                     /replace="c"
     conflict        5310..5311
                     /gene="splA"
                     /note="reported as GT in mRNA sequence, GenBank Accession
                     Number M33785"
                     /citation=[1]
                     /replace="gt"
BASE COUNT     2247 a   1037 c    857 g   1790 t
ORIGIN

Query Match         7.9%;      Score 49;   DB 3;    Length 5931;
Best Local Similarity  47.3%;    Pred. No. 0.2;   Gaps    0;
Matches    148;   Conservative    0;   Mismatches   165;   Indels   0;   Gaps  0;

Qy     27  caacagtaagtacttcaaaacaaaaacatcatggcaacagacagtgggagtacttgaataa  86
           |||||| ||  ||||||||| ||| |||| |||||| || || ||| || || || ||
Db    722  CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATACAA  781

Qy     87  caacaatggcgacaataacaacacacaatatgttacaagctcattacgagataattacg  146
           || ||| |||| || || ||| || || ||| || ||| || || ||| || || |||
Db    782  CAATAATAATAACAATAACAACAACAACAACAACAACCATTACCACCACAAAAAGAGG  841

Qy    147  aaatgaagatcattacttggtggactattccttggcctccaagatccttatacatgtag  206
           ||| || || || || ||| || |||| || || || || ||| || || || || ||
Db    842  AGATTCTAGTGGGTTAATATGAGCTTTAAATTCTTTAAAAGAAAAAATTGGTCCCATCCAA  901
```

```
Db   971  ATTTAGTTGGTGATCTATTAAATTAAAGTAGTAATAATAATGGTAGAACTA 1030
Qy   254  tgaatgttcatagagagatagagccatttgagacaatcaccacctagagatattaata 313
Db  1031  GTACACTTAATTATTATTAAATGGTACTTTAATAATAAAAATATTTATGAAGTAATATTA 1090
Qy   314  ggtattcttcttctaaacctaatctgaccaaaacctaacttaccctagtcataacc 373
Db  1091  ATAATAATTAATAATAAATTAAATTATATCCCTTCTAACATATTTATATAATTCTAATA 1150
Qy   374  ctagtttttcaagaaaat 391
Db  1151  TTAATAAAAATGGTAAAT 1168
```

RESULT 8
YSCMTVAR1

```
LOCUS       YSCMTVAR1    1200 bp    DNA            PLN    30-SEP-1988
DEFINITION  Yeast (S.cerevisiae) mitochondrial var1 gene, 40.0 allele.
ACCESSION   J01525
NID         g343948
KEYWORDS    repeat region; unidentified reading frame.
SOURCE      Yeast (S.cerevisiae) DNA.
ORGANISM    Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1 (bases 1 to 1200)
AUTHORS     Hudspeth,M.E.S., Ainley,W.M., Shumard,D.S., Butow,R.A. and
            Grossman,L.I.
TITLE       Location and structure of the var1 gene one yeast mitochondrial
            DNA: Nucleotide sequence of the 40.0 allele
JOURNAL     Cell 30, 617-626 (1982)
MEDLINE     83050946
COMMENT     Alleles of the var1 locus on yeast mitochondrial DNA specify the
            size of var1 ribosomal RNA. [1]: given sequence determines the
            smallest var1 protein. The coding region is highly unusual: 89.6%
            at with a 46 bp 'gc'-rich palindromic cluster that accounts for 38%
            of the total 'gc' residues. results strongly suggest that like
            mammalian mitochondria but unlike those from Neurospora, yeast
            mitochondria use 'aua' as a methionine codon. Comparison with
            sequence of a var1 allele specifying a larger protein suggests that
            some size polymorphism of var1 protein results from in-frame
            insertions of a variable number of 'aat' (asn) codons. [1] compares
            sequence with that of the 42.0 allele of ds401.
FEATURES    Location/Qualifiers
   source   1..1200
            /organism="Saccharomyces cerevisiae"
            /mitochondrion
            /db_xref="taxon:4932"
   CDS      10..1200
            /note="var1(40.0)"
            /codon_start=1
            /transl_table=3
            /db_xref="PID:g343949"
            /translation="MKLKLNIILSIINKTNNNNNIIINNTLDSLINKKLLLKNILLD
            INNKKINNIKRILNNNINPAGANPVVHRIGPAGNINNKIISKTINQHSLNKLNIKFYYNNDI
            NIEIINTINDKLINKLLYKIITLKLNNININKIIISKTINQHSLNKLNIKFYYNNDI
            NNNNNNNYININLININNINNNLCNILSYYYNKKVTIEPIKLSYIYLNSDIF
            SKYISLNDIDKYNNGILTNYQRLNNIMPKLNDHNISINYINNINNNKYNNINL
            LNNNNNINNNNNYYNNNNNYIGNINNIYNNITIDNIPIDILIYKYLVGKSIKFKGRLS
            NNNGRTSTTNLLGTFNNKKYLXSNINNNYKLNYIPSNHNLYNNSNINKNGKYNIKVK
            LNFI"
BASE COUNT       605 a      57 c      67 g     471 t
ORIGIN      About 125 bp upstream of HinfI site.

Query Match                 7.6%;    Score 46.8;  DB 8;  Length 1200;
Best Local Similarity   45.2%;   Pred. No. 0.74;
Matches  171;  Conservative  0;  Mismatches 207;  Indels  0;  Gaps  0;

Qy   14  atagaaacagttcaacagtaagtacttcaaaaacaaaagcatccatggcaagacagatgg 73
```

```
Qy   207  ctttgtaaagggaattagatctgctcaagctcttggtggaccacatgaatgttcatag 266
Db   902  TGAATATAGATGGGATTAAGAAAATCAAATTCACTTACATTAAATATTGAAGATAAAG 961
Qy   267  aagagataggagcccatttgagacaatcaccacctagagatattaataggtattcctcct 326
Db   962  TAGAGTGTTCTTACAGACTACCAACACATCTCGTTCAAAAGGTGTAAGTATTTTTTCTTTTTTT 1021
Qy   327  aaaccttaatctt 339
Db  1022  AAAATTTTATTTT 1034
```

RESULT 7
MISCVA

```
LOCUS       MISCVA       1200 bp    DNA            PLN    01-OCT-1996
DEFINITION  Yeast mitochondrial var1 gene (allele 40.0) a ribosomal protein.
ACCESSION   V00705 J01525
NID         g13610
KEYWORDS    ribosomal protein.
SOURCE      baker's yeast.
ORGANISM    Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1 (bases 1 to 1200)
AUTHORS     Hudspeth,M.E., Ainley,W.M., Shumard,D.S., Butow,R.A. and
            Grossman,L.I.
TITLE       Location and structure of the var1 gene on yeast mitochondrial DNA:
            nucleotide sequence of the 40.0 allele
JOURNAL     Cell 30 (2), 617-626 (1982)
MEDLINE     83050946
FEATURES    Location/Qualifiers
   source   1..1200
            /organism="Saccharomyces cerevisiae"
            /mitochondrion
            /db_xref="taxon:4932"
   CDS      10..1200
            /note="var1 reading frame"
            /codon_start=1
            /transl_table=3
            /db_xref="PID:g13611"
            /db_xref="SWISS-PROT:P02381"
            /translation="MKLKLNMILSMMNKTNNNNNIIINNTLDSLMNKKLLLKNMLLD
            MNNKKMNNMKRMLNNNMNPAGANPVVHRIGPAGNINNKLQHLNMNNWNTQIYNYNK
            NMEIMNTMNDKLINKLLYKMMTLKLNNNNIIMSKTINQHSLNKLNIKFYYNNDI
            NNNNNNNYYMNMNMNKLMINMNMNNNLCNILSYYYNKKVTIEPIKLSYIYLNSDIF
            SKYISLNDKYNNGILTNYQRMLNNINNNYIGNINNIYNNITIDNIPIDILMYKYLVGKSIKFKGRLS
            NNNGRTSTTNLLGTFNNKKYLWSNINNNYKLNYIPSNHNLYNNSNINKNGKYNIKVK
            LNFI"
BASE COUNT       605 a      57 c      67 g     471 t
ORIGIN

Query Match                 7.6%;    Score 46.8;  DB 8;  Length 1200;
Best Local Similarity   45.2%;   Pred. No. 0.74;
Matches  171;  Conservative  0;  Mismatches 207;  Indels  0;  Gaps  0;

Qy   14  atagaaacagttcaacagtaagtacttcaaaaacaaaagcatccatggcaagacagatgg 73
Db  791  ATATTAATAATAATAATAATAATAATAATAATAATATTAATTATTAAATATAAATATAAATA 850
Qy   74  agtacttgaataacaacaatggcgacaataacaacaacaataagttacaagctcattac 133
Db  851  ATAATATTAATAATAATAATAATAATAATAATAATAATAATAATTGGTAATATTA 910
Qy  134  gagataattatggaatgaagatcatttacttggtggactattcctcttggcctccaagat 193
Db  911  ATAATATTTATAATAATAATAACTATTGATAATATTCCTATAGATATTTTAATAATATAAAT 970
Qy  194  cttatacatgtagctttgtaaaagggaattagatctgctcaagctcttggtggacaca 253
```

```
Db   791 ATATTAATAATTATTAATAATAAATATAATAAATATATAAATATTAATTTATTATTAAAATAATAATATA 850
```

```
Qy    14 atagaaacagttcaacagtagtagtacttcaaaaacaaaagcatcatgcgcaagacagagatgg 73
         ||| |||||||||||||||||||||||||  ||||||||||||||||| |
Db   791 ATATTAATAATTATTAATAATAAATATAATAAATATATAAATATTAATTTATTATTAAAATAATAATA 850

Qy    74 agtacctgaataacaacaatggcgacaataacaacaataatgttacaagctcattac 133
         |--|||||||  |||||| |||  |||||| |||||||  |
Db   851 ATAATATTAATAATTATTAATAATAAATATAATAAATATATAAATATTAATTTATTATTA 910

Qy   134 gagataattggaaatgaagatcattcacttggtggactattcctcttggcctccaagat 193
         ||| |||||  | |||||| |||  |||||| || || ||||||||
Db   911 ATAATATTTATAATTATTAATAATAAATATGATAATACTATTGATATATTTTTAATATATAAAT 970

Qy   194 cttatacatgtagctttgtaaaaggggaattagatctgctcaagctcttggtggacaca 253
         ||-|| ||||| ||||||| ||  || |||||||
Db   971 ATTTAGTTGGTTGATCTATTAAATTTAAAAGGTAGATTTAAAGTAATAAGTGTAGAACTA 1030

Qy   254 tgaatgttcatagaagagagagcatttgagacaatcaccacctagagatattaata 313
         ||-|||||||||| || |||| |||||| |||| |
Db  1031 GTACACTTAATTTATTAAATGTACTTTTAATAATAAAAAAATATTTATGAAGTAATATTA 1090

Qy   314 ggtatctctctaaacctaatctgaaccaaacctaacttttaccctagtcataaacc 373
         ||-|||||||  |||||||| |||||||||||||||| ||||||
Db  1091 ATAATATTATAAATTAAATTATATCCCTTCTAATCATATAATTTATATAATAATTCTAATA 1150

Qy   374 ctagtttttcaagaaat 391
         ||-|||| |||||
Db  1151 TTAATAAAAATGGTAAAT 1168
```

```
RESULT  9
MISCVA
LOCUS       MISCVA      1200 bp    DNA            PLN       01-OCT-1996
DEFINITION  Yeast mitochondrial var1 gene (allele 40.0) a ribosomal protein.
ACCESSION   V00705 J01525
NID         g13610
KEYWORDS    ribosomal protein.
SOURCE      baker's yeast.
ORGANISM    Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 1200)
AUTHORS     Hudspeth,M.E., Ainley,W.M., Shumard,D.S., Butow,R.A. and
            Grossman,L.I.
TITLE       Location and structure of the var1 gene on yeast mitochondrial DNA:
            nucleotide sequence of the 40.0 allele
JOURNAL     Cell 30 (2), 617-626 (1982)
MEDLINE     83050946
FEATURES             Location/Qualifiers
     source          1..1200
                     /organism="Saccharomyces cerevisiae"
                     /mitochondrion
                     /db_xref="taxon:4932"
     CDS             10..1200
                     /note="var1 reading frame"
                     /codon_start=1
                     /transl_table=3
                     /db_xref="PID:g13611"
                     /db_xref="SWISS-PROT:P02381"
                     /translation="MKLKLLNMILSMNKTNNNNNNIINNTLDSLMNKKLLLKNMLLD
                     MNNKKMNNMKRMLNNNNNPAGANPVVHRIGPAGNINNKLQHLNMNMNWNTQIYNYNK
                     NMEIMTMNDKLINKLLYKMMTLKLNMMNINKIIMSKTINQHSLNKLNIKFYYYNNDI
                     NNNNNNNYMNMMNKLMNIMNNNNNNNCNILSYYNKKVTIEPIKLSYIYLNSDIF
                     SKYISLNDMDKYNNGILTNYQRMLNIMPKLNDHNISMNYINIININNNKYNNMINL
                     LNNNNINNNNNYNNNNNNYIGNINNIYNNMTIDNIPMDILMYKYLVGWSIKFKGRLS
                     NNNGRTSTTNLLNGTFNNKKYLWSNINNNYKLNYIPSNHNLYNNSNINKNGKYNIKVK
                     LNFI"
BASE COUNT      605 a    57 c    67 g    471 t
ORIGIN
```

```
Db   791 ATATTAATAATTATTAATAATAAATATAATAAATATTAATTTATTAAAATAATAATATA 850
```

```
Qy    14 atagaaacagttcaacagtagtagtacttcaaaaacaaaagcatcatgcgcaagacagagatgg 73
         ||| |||||||||||||||||||||||||  ||||||||||||||||| |
Db   791 ATATTAATAATTATTAATAATAAATATAATAAATATATAAATATTAATTTATTATTAAAATAATAATA 850

Qy    74 agtacctgaataacaacaatggcgacaataacaacaataatgttacaagctcattac 133
         |--|||||||  |||||| |||  |||||| |||||||  |
Db   851 ATAATATTAATAATTATTAATAATAAATATAATAAATATATAAATATTAATTTATTATTATTGGTAATATTA 910

Qy   134 gagataattggaaatgaagatcattcacttggtggactattcctcttggcctccaagat 193
         ||| |||||  | |||||| |||  |||||| || || ||||||||
Db   911 ATAATATTTATAATTATTAATAATAAATATGATAATACTATTGATATATTTTTAATATATAAAT 970

Qy   194 cttatacatgtagctttgtaaaaggggaattagatctgctcaagctcttggtggacaca 253
         ||-|| ||||| ||||||| ||  || |||||||
Db   971 ATTTAGTTGGTTGATCTATTAAATTTAAAAGGTAGATTTAAAGTAATAAGTGTAGAACTA 1030

Qy   254 tgaatgttcatagaagagagagcatttgagacaatcaccacctagagatattaata 313
         ||-|||||||||| || |||| |||||| |||| |
Db  1031 GTACACTTAATTTATTAAATGTACTTTTAATAATAAAAAAATATTTATGAAGTAATATTA 1090

Qy   314 ggtatctctctaaacctaatctgaaccaaacctaacttttaccctagtcataaacc 373
         ||-|||||||  |||||||| |||||||||||||||| ||||||
Db  1091 ATAATATTATAAATTAAATTATATCCCTTCTAATCATATAATTTATATAATAATTCTAATA 1150

Qy   374 ctagtttttcaagaaat 391
         ||-|||| |||||
Db  1151 TTAATAAAAATGGTAAAT 1168
```

```
RESULT 10
YSCMTVAR1
LOCUS       YSCMTVAR1   1200 bp    DNA            PLN       30-SEP-1988
DEFINITION  Yeast (S.cerevisiae) mitochondrial var1 gene, 40.0 allele.
ACCESSION   J01525
NID         g343948
KEYWORDS    repeat region; unidentified reading frame.
SOURCE      Yeast (S.cerevisiae) DNA.
ORGANISM    Mitochondrion Saccharomyces cerevisiae
            Eukaryotae; mitochondrial eukaryotes; Fungi; Ascomycota;
            Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
            Saccharomyces.
REFERENCE   1  (bases 1 to 1200)
AUTHORS     Hudspeth,M.E.S., Ainley,W.M., Shumard,D.S., Butow,R.A. and
            Grossman,L.I.
TITLE       Location and structure of the var1 gene one yeast mitochondrial
            DNA: Nucleotide sequence of the 40.0 allele
JOURNAL     Cell 30, 617-626 (1982)
MEDLINE     83050946
COMMENT     Alleles of the var1 locus on yeast mitochondrial DNA specify the
            size of var1 ribosomal RNA. [1]: given sequence determines the
            smallest var1 protein. The coding region is highly unusual: 89.6%
            at with a 46 bp 'gc'-rich palindromic cluster that accounts for 38%
            of the total 'gc' residues. results strongly suggest that like
            mammalian mitochondria but unlike those from Neurospora, yeast
            mitochondria use 'aua' as a methionine codon. Comparison with
            sequence of a var1 allele specifying a larger protein suggests that
            some size polymorphism of var1 protein results from in-frame
            insertions of a variable number of 'aat' (asn) codons. [1] compares
            sequence with that of the 42.0 allele of ds401.
FEATURES             Location/Qualifiers
     source          1..1200
                     /organism="Saccharomyces cerevisiae"
                     /mitochondrion
                     /db_xref="taxon:4932"
     CDS             10..1200
                     /note="var1(40.0)"
                     /codon_start=1
                     /transl_table=3
                     /db_xref="PID:g343949"
                     /translation="MKLKLLNILSLINKTNNNNNIIINNTLDSLINKKLLLKNILLD
                     INNKINNIKRILNNNINPAGANPVVHRIGPAGNINNKLQHLNMNNXNTQIYNYNK
                     NIEINTINDKLINKLLYKIITLKLNNINNKIIISKTINQHSLNKLNIKFYYYNNDI
                     NNNNNNNYYINIINKLINIININNNINNNLCNILSYYNKKVTIEPIKLSYIYLNSDIF
```

```
Query Match             7.6%;  Score 46.8;  DB 38;  Length 1200;
Best Local Similarity  45.2%;  Pred. No. 0.74;
Matches  171;  Conservative    0;  Mismatches 207;  Indels    0;  Gaps    0;
```

```
                    SKYISLNDIDKYNNGILTNYQRILNNIMPKLNDHNISINYINNNINNNKYNNIINL
                    LNNNNINNNNNYNNNNNNNYIGNINNIYNNITIDNIPIDILIYKYLVGXSIKFKGRLS
                    NNNGRTSTNLLNGTFNNKKYLXSNINNVKLNYIPSNHHNLYNNSNINKNGKYNIKVK
                    LNFI"

source          1 (sites)
BASE COUNT      605 a    57 c    67 g    471 t
ORIGIN          About 125 bp upstream of HinfI site.

Query Match         7.6%;   Score 46.8;  DB 38;  Length 1200;
Best Local Similarity   45.2%;   Pred. No. 0.74;
Matches  171;  Conservative   0;  Mismatches 207;  Indels   0;  Gaps   0;

Qy    14 atagaaacagttcaacagtagtacctcaaaaacaaaagcatcatgcaagacagatgg  73
         |||  ||||  ||  ||  || | | |    |   ||  ||  |  | |  |  ||
Db   791 ATATTAATAATATTAATAATAATAAATAATAATAATTATTAATAATTAATAATAATA 850

Qy    74 agtacttgataacaacaatggcgacaataacaacaacaataatgttacaagctcattac 133
         || |||||| ||| |||||||| ||||| | |||| |   ||||   ||  ||  |
Db   851 ATAATATTAATAATATTAATAATAATAATATTAATAATAATATTATTGTAATATTA 910

Qy   134 gagataattggaaatgaagatcatttacttggtggactattcctcttggcctccaagat 193
         |||| |||| |  | |||| |  ||||||   ||| || ||   || |||||||
Db   911 ATAAATATTTAATAATATTAATAACTATTGATAATATTCCTATAGATATTTAATATATAAAT 970

Qy   194 cttatacatgtagctttgtaaaaggaatttagatctgctcaagctcttggtggacaca 253
         || |||||||||  || ||  |||||||||||||| |||||| |||||||  ||| |
Db   971 ATTTAGTTGGTTGATCTATTAAATTTAAAGGTAGATTAAGTAATAATAATGGTAGAACTA 1030

Qy   254 tgaagttcatagaagagagagagccatttgaccaataccaccctagaagatattaata 313
         | |||| ||| |||||| ||||||||||  ||| |   |||| ||| || ||||||
Db  1031 GTACACTAATTTAATTAATAAATGGTACTTTTAATAATAAAAAATATTTATGAAGTAACTATTA 1090

Qy   314 ggtattcctctctaaaccctaatcctgaaccaaaaccctaactttacccctagtcataacc 373
         ||| |||| ||| |||||||   |||| ||||| |||||   ||  ||  |||||  ||
Db  1091 ATAAATATTATAAAAATTAAATTAATATATCCCTTCTAATCATAATTTATATAAATAATTCTAATA 1150

Qy   374 ctagttttcaagaaaat 391
         ||  |||| ||||
Db  1151 TTAATAAAATGGTAAAT 1168
```

RESULT 11
PETZFDB2
LOCUS   PETZFDB2   1083 bp   mRNA   PLN   10-MAR-1995
DEFINITION   Petunia mRNA for zinc-finger DNA binding protein.
ACCESSION   D26084
NID   g439488
KEYWORDS   zinc-finger DNA binding protein.
SOURCE   Petunia hybrida (strain:Titan parade) cDNA to mRNA.
ORGANISM   Petunia x hybrida
               Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
               Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
               plants; Magnoliophyta; Magnoliopsida; Solananae; Solanales;
               Solanaceae; Petunia.
REFERENCE   1 (sites)
AUTHORS   Takatsuji.H., Nakamura,N. and Katsumoto.Y.
TITLE   A new family of zinc finger proteins in petunia: structure, DNA
               sequence recognition, and floral organ-specific expression
JOURNAL   Plant Cell 6 (7), 947-958 (1994)
MEDLINE   94348284
REFERENCE   2 (bases 1 to 1083)
AUTHORS   Takatsuji.H.
JOURNAL   Unpublished (1995)
COMMENT   Submitted (06-Dec-1993) to DDBJ by:
               Hiroshi Takatsuji
               Department of Applied Physiology
               National Institute of Agrobiological Resources
               2-1-2 kannondai, Tsukuba
               Ibaraki 305
               Japan
               Phone: 0298-38-8383
               Fax: 0298-38-7417.
FEATURES   Location/Qualifiers

---

source   1 (sites)
                /organism="Petunia x hybrida"
                /strain="Titan parade"
                /db_xref="taxon:4102"
CDS   109..933
                /codon_start=1
                /product="zinc-finger DNA binding protein"
                /db_xref="PID:d1005618"
                /db_xref="PID:g439489"
                /translation="MALEALKSPTAATPSLPPRYEDHVDMNLDSWVGKRSKRPRIE
                TPPSEEEYLALCLIMLARSGNGTTPGSTDTTTTTISKEPBKKNRELTPVHOETEQSY
                KCSVCDKSFSSYQALGHGKASHRKITTIATTALLDDNNNNPTTSNSTSGNVVNNISAL
                NPSGRSHVCSICHKAFPTGQALGGHKRRHYEGKLGGGNNNHRDGGGHGSGSVVTTSDGG
                ASTHTLRDFDLNMLPPSPELQLGLSIDCDLKSQIPIEQEVESPMPLKKPRLLFSMD"

BASE COUNT   328 a   216 c   198 g   341 t
ORIGIN

Query Match   7.5%;   Score 46.2; DB 8; Length 1083;
Best Local Similarity   69.2%;   Pred. No. 1;
Matches 63; Conservative   0; Mismatches 28; Indels   0; Gaps   0;

Qy   192 atcttatacatgtagctttgtaaaaggaattagatctgctcaagctcttggtggaca 251
         ||||| ||| ||||| ||| | |||||||| |||| | ||||||||||||||||
Db   408 ATCTTACAAGTGTAGCGTGTGTGACAAGTCTTTTAGTTCTTATCAAGCTCTTGGTGGACA 467

Qy   252 catgaatgttcatagaagatagagccatt 282
         | | ||||| || ||||||| ||||||
Db   468 CAAGCAAGTCATAGGAAAATTACAACTATT 498

RESULT 12
PETZFDB2
LOCUS   PETZFDB2   1083 bp   mRNA   PLN   10-MAR-1995
DEFINITION   Petunia mRNA for zinc-finger DNA binding protein.
ACCESSION   D26084
NID   g439488
KEYWORDS   zinc-finger DNA binding protein.
SOURCE   Petunia hybrida (strain:Titan parade) cDNA to mRNA.
ORGANISM   Petunia x hybrida
               Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
               Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
               plants; Magnoliophyta; Magnoliopsida; Solananae; Solanales;
               Solanaceae; Petunia.
REFERENCE   1 (sites)
AUTHORS   Takatsuji.H., Nakamura,N. and Katsumoto.Y.
TITLE   A new family of zinc finger proteins in petunia: structure, DNA
               sequence recognition, and floral organ-specific expression
JOURNAL   Plant Cell 6 (7), 947-958 (1994)
MEDLINE   94348284
REFERENCE   2 (bases 1 to 1083)
AUTHORS   Takatsuji.H.
JOURNAL   Unpublished (1995)
COMMENT   Submitted (06-Dec-1993) to DDBJ by:
               Hiroshi Takatsuji
               Department of Applied Physiology
               National Institute of Agrobiological Resources
               2-1-2 kannondai, Tsukuba
               Ibaraki 305
               Japan
               Phone: 0298-38-8383
               Fax: 0298-38-7417.
FEATURES   Location/Qualifiers
source   1..1083
                /organism="Petunia x hybrida"
                /strain="Titan parade"
                /db_xref="taxon:4102"
CDS   109..933
                /codon_start=1
                /product="zinc-finger DNA binding protein"
                /db_xref="PID:d1005618"
                /db_xref="PID:g439489"
                /translation="MALEALKSPTAATPSLPPRYEDHVDMNLDSWVGKRSKRPRIE
```

```
                    TPPSEEYLALCLIMLARSGNGTPGSTDTTTTISKEPEKKNRELPVHQETEQSY
                    KCSVCDKSFSSYQALGGHKASHRKITTIATTALLDDNNNNPTTSNSTGNVVNISAL
                    NPSGRSHVCSICHKAFPTGQALGGHKRRHYEGKLGGNNNNHRDGGGHSGSVVTTSDGG
                    ASTHTLRDFDLNMLPPSPELQLGLSIDCDLKSQPIEQEVESPMPLKKPRLLFSMD"

BASE COUNT     328 a    216 c    198 g    341 t
ORIGIN

Query Match
Best Local Similarity   7.5%;   Pred. No. 1;                       Score 46.2;  DB 38;  Length 1083;
Matches  63;   Conservative  0;  Mismatches  28;  Indels  0;  Gaps  0:

Qy    192 atcttatacatgtagctttgtaaaagggaattagatctgctcaagctcttggtggaca 251
          ||||||  ||||| |||  ||||| |||  |||| ||||||||||| ||||| |||||
Db    408 ATCTTACAAGTGTAGCGTGTGACAAGTCTTTTAGTTCTTATCAAGCTCTTGGTGGACA 467

Qy    252 catgaatgttcatagaaggagatagagccatt 282
          || ||||||||||||||| || | |||
Db    468 CAAAGCAAGTCATAGGAAAATTACAACTATT 498

RESULT 13
PETZFDB1
LOCUS       PETZFDB1       3041 bp       DNA            PLN      10-MAR-1995
DEFINITION  Petunia zinc-finger DNA binding protein gene.
ACCESSION   D26083
NID         g439486
KEYWORDS    zinc-finger DNA binding protein.
SOURCE      Petunia hybrida (strain:Titan parade) DNA.
ORGANISM    Petunia x hybrida
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
            plants; Magnoliophyta; Magnoliopsida; Solananae; Solanales;
            Solanaceae; Petunia.
REFERENCE   1  (sites)
AUTHORS     Takatsuji,H., Nakamura,N. and Katsumoto,Y.
TITLE       A new family of zinc finger proteins in petunia: structure, DNA
            sequence recognition, and floral organ-specific expression
JOURNAL     Plant Cell 6 (7), 947-958 (1994)
MEDLINE     94348284
REFERENCE   2  (bases 1 to 3041)
AUTHORS     Takatsuji,H.
JOURNAL     Unpublished (1995)
COMMENT     Submitted (06-Dec-1993) to DDBJ by:
            Hiroshi Takatsuji
            Department of Applied Physiology
            National Institute of Agrobiological Resources
            2-1-2 kannondai, Tsukuba
            Ibaraki 305
            Japan
            Phone:  0298-38-8383
            Fax:    0298-38-7417.
FEATURES             Location/Qualifiers
     source          1..3041
                     /organism="Petunia x hybrida"
                     /strain="Titan parade"
                     /db_xref="taxon:4102"
     misc_feature    1348..1357
                     /note="MADS box protein binding sequence"
     misc_feature    1591..1600
                     /note="MADS box protein binding sequence"
     TATA_signal     1913..1924
     CDS             2057..2890
                     /codon_start=1
                     /product="zinc-finger DNA binding protein"
                     /db_xref="PID:d1005617"
                     /db_xref="PID:g439487"
                     /translation="MALEALKSPTAATPLPPRYEDQVDMSNLDSWVGKRSKRPRIE
                     TPPSEEYLALCLIMLARSGNGTPPSSIPGSTDTTTISKEPEKKNRDVAPVYQETEQS
                     YKCSVCDKSFSSYQALGGHKASHRKITTIATTALLDDNNNNPTTSNSTNGNVVNNIST
                     LNPSGRSHVCSICHKAFPSGQALGGHKRRHYEGKLGGNNNNHRDGGGHSGSVVTTSD
                     GGASTHTLRDFDLNMLPPSPELQLGLSIDCGLKSQQVPIEQEVESPMPLKKPRLLFSM
                     D"
```

```
BASE COUNT     977 a    552 c    470 g   1042 t
ORIGIN

Query Match
Best Local Similarity   7.2%;   Pred. No. 1.9;                     Score 44.6;  DB 8;  Length 3041;
Matches  62;   Conservative  0;  Mismatches  29;  Indels  0;  Gaps  0:

Qy    192 atcttatacatgtagctttgtaaaagggaattagatctgctcaagctcttggtggaca 251
          ||||||  ||||| |||  ||||| |||  |||| ||||||||||| |||||
Db   2359 ATCTTACAAGTGTAGCGTGTGACAAGTCTTTTAGTTCTTATCAAGCTCTTGGCGGACA 2418

Qy    252 catgaatgttcatagaaggagatagagccatt 282
          || ||||||||||||||| || | |||
Db   2419 CAAAGCAAGTCATAGGAAAATTACAACTATT 2449

RESULT 14
PETZFDB1
LOCUS       PETZFDB1       3041 bp       DNA            PLN      10-MAR-1995
DEFINITION  Petunia zinc-finger DNA binding protein gene.
ACCESSION   D26083
NID         g439486
KEYWORDS    zinc-finger DNA binding protein.
SOURCE      Petunia hybrida (strain:Titan parade) DNA.
ORGANISM    Petunia x hybrida
            Eukaryotae; mitochondrial eukaryotes; Viridiplantae;
            Charophyta/Embryophyta group; Embryophyta; vascular plants; seed
            plants; Magnoliophyta; Magnoliopsida; Solananae; Solanales;
            Solanaceae; Petunia.
REFERENCE   1  (sites)
AUTHORS     Takatsuji,H., Nakamura,N. and Katsumoto,Y.
TITLE       A new family of zinc finger proteins in petunia: structure, DNA
            sequence recognition, and floral organ-specific expression
JOURNAL     Plant Cell 6 (7), 947-958 (1994)
MEDLINE     94348284
REFERENCE   2  (bases 1 to 3041)
AUTHORS     Takatsuji,H.
JOURNAL     Unpublished (1995)
COMMENT     Submitted (06-Dec-1993) to DDBJ by:
            Hiroshi Takatsuji
            Department of Applied Physiology
            National Institute of Agrobiological Resources
            2-1-2 kannondai, Tsukuba
            Ibaraki 305
            Japan
            Phone:  0298-38-8383
            Fax:    0298-38-7417.
FEATURES             Location/Qualifiers
     source          1..3041
                     /organism="Petunia x hybrida"
                     /strain="Titan parade"
                     /db_xref="taxon:4102"
     misc_feature    1348..1357
                     /note="MADS box protein binding sequence"
     misc_feature    1591..1600
                     /note="MADS box protein binding sequence"
     TATA_signal     1913..1924
     CDS             2057..2890
                     /codon_start=1
                     /product="zinc-finger DNA binding protein"
                     /db_xref="PID:d1005617"
                     /db_xref="PID:g439487"
                     /translation="MALEALKSPTAATPLPPRYEDQVDMSNLDSWVGKRSKRPRIE
                     TPPSEEYLALCLIMLARSGNGTPPSSIPGSTDTTTISKEPEKKNRDVAPVYQETEQS
                     YKCSVCDKSFSSYQALGGHKASHRKITTIATTALLDDNNNNPTTSNSTNGNVVNNIST
                     LNPSGRSHVCSICHKAFPSGQALGGHKRRHYEGKLGGNNNNHRDGGGHSGSVVTTSD
                     GGASTHTLRDFDLNMLPPSPELQLGLSIDCGLKSQQVPIEQEVESPMPLKKPRLLFSM
                     D"

BASE COUNT     977 a    552 c    470 g   1042 t
ORIGIN
```

Query Match
Best Local Similarity   7.2%;   Score 44.6;   DB 38;   Length 3041;
Matches   62;   Conservative   0;   Mismatches   29;   Indels   0;   Gaps   0;

```
Qy   192 atcttatacatgctagctttgtaaaagggaattagatctgctcaagctcttggtggaca 251
         |||||||| ||||| || ||| |||| |||||||||||||||||||||||  |||||
Db  2359 ACTTACAAGTGTAGCGTGTGCACAAGTCTTTTAGTTCTTATCAAGCTCTTGCGGACA 2418

Qy   252 catgaatgttcatagagagatagagagccatt 282
         || |||||||| ||||| | | | || ||
Db  2419 CAAAGCAAGTCATAGGAAAATACAACTATT 2449
```

RESULT 15
AF079445
LOCUS   AF079445   12668 bp   DNA   INV   16-AUG-1998
DEFINITION Dictyostelium discoideum TipC (tipC) gene, complete cds.
ACCESSION   AF079445
NID   g3420744
KEYWORDS   .
SOURCE   Dictyostelium discoideum.
ORGANISM   Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 12668)
AUTHORS   Stege,J.T. and Loomis,W.F.
TITLE   Interaction of tip genes in early Dictyostelium discoideum
            development
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 12668)
AUTHORS   Stege,J.T., Laub,M.T. and Loomis,W.F.
TITLE   Direct Submission
JOURNAL   Submitted (21-JUL-1998) Biology, UCSD, 9500 Gilman Drive, La Jolla,
            CA 92093, USA
FEATURES   Location/Qualifiers
     source   1..12668
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
     mRNA   join(<227..442,522..>11852)
                     /gene="tipC"
                     /product="tipC"
     gene   <227..>11852
                     /gene="tipC"
     CDS   join(227..442,522..11852)
                     /gene="tipC"
                     /note="involved in early development; necessary for tip
                     formation; similar to Saccharomyces cerevisiae vacuolar
                     protein sorting-associated protein VPS13 encoded by
                     sequence in GenBank Accession Number AF001317; similar to
                     Schizosaccharomyces pombe vacuolar protein
                     sorting-associated protein encoded by sequence in GenBank
                     Accession Number AL021767; similar to Caenorhabditis
                     elegans T08G11.1 sequence in GenBank Accession Number
                     280220"
                     /codon_start=1
                     /product="tipC"
                     /db_xref="PID:g3420745"
                     /translation="MVSHIAASVHTKYLGDYVDDLNKDNIKLSFLSGEAVLQDLKIKK
                     TVLQSFLPNVIVKQAIIKKLSSLHVPWDLKGKPAIIKIEGIYVLAETSVEFDEQYKK
                     KFQDERQAKLHIQEVLRIKHKQQLKNPHQSTTTTTTTSDESNTFGSKLLQTVDNLQLY
                     IDSVHIRFEDSVNRRSFSFGVTLNSLVAESTDTWNPTFIKNESTIIHKLINLNQLSI
                     YWNSNSPKLKTYNIDDLSQKLKSMIKKEDSGKQQQQQQQGEEQDDEIEEDYFLSTE
                     SFKLFEKSIEFRSARPINTTIKQSPRKWNYAIQCVLQKVHQKRYTQSWSYIKEFLSD
                     KRDYIELFKKLKKTIHQTEQTRLDSLEWKLFPDQILFFRNLSFKIIEKEKKLEKERL
                     KLQQQQQQQQQNLNKISPTTTTTPSTSTGGGGWLSSWWKPIVNTSVPITAATTTTAT
                     NTDSTITKLKDIKLSKDDWNEIYDTIGYDENDDIINSNVASLQQPKOFGKDFSNIIKT
                     VINFRLVQGGLKLKKKRTIALLQLNDISLSLKSKINSSFVFNGNLKGLELIDQSTFN
                     TQFPKLISPLLKRNSSIPTKIISQPPIFNITLESHSLSSTITSSTTTATTTQSTSTN
                     SNNNNNIKVNYSLSIQSKPLTIVYYPKFISTVNDFFFKKEDDDVQDMVLEDLEKKAQ
                     ETIESIKAQTRDKLLLAISNKTTFSIDMDLEAPIILVPESITNKNTNLLILDLGKLLJ
                     NHYPKQQQLQQQLQQLQEQGEQGEOQEQKQEQEQKQEQEQKDWKEEFYDEYLFKLINTQV
                     LLANLKNDWRDIEQVKKFRMNIANPLNANFTLKISKLNNRQLLTSKLMASIDLLEFY
```

LSSCQYVNLVNIIQSVTQPIDQQQLIGESMYMSAFDYSSVYQSNIFKPQQLQPQQ
KQQSPPLISSPPLIKKKSYSKEFIDSKVMFEASFNIDKFNIHLRLDNYDSSGSSSNN
DSSNTTTNHNQLVSHNVALVYLRCLHGTYVQKLFDTNLELLIKGMWIEDCFQKPQSI
AHGSGGDYLATTTNRFKPNDVNIGGSNENLSILKNNQNNNQNNNQNNNESSPTVFITSPP
QINLILNRKTVAGLIEFSNSVSNLSILKNNQNNNQNNNQNNNQNNNESSPTVFITSPP
PPPPPLNHQSNNNSQYDTLSSSTYLIRNSNNKILPPIQQISTTANNNIKIQTRATIK
IDLVRILLTRENNNPLIKASFTGFNCTTDQYNYKTIMNGKLGSKLYDMTEGRNYRT
ILTRERKLNNKTSPTTPSSSGVSTGGTFDIDIDNFDDVSSSSSSSSLLSSSSLLISPP
GGEQSSLVFQFESRDNNSQMLKINLSSIRFFLKRFTEELRLFLNNVNMREYLKS
SIYSAATVISQNRSTLFYEIEIQNPYIVPLSSLSNKIFIVDLGKINIKNQFEKIIPT
STEIDMLMEPVIIENIIIEAQNIKLLSGLNSEIKKSENNNNNNSNTNQRSYGTM
VSDVNVSINLKTPLFSELNLKQYLEQCMKIQSWEKHFTISQFELNVSEYELKSLIELL
DGNLSEFSSEILNDANSSFDQKQQEIEKLKLQQEIEDNNDNEDEQVNKLLQLIKEQ
VGVTYCKLGKFSIMFMKRDGTDIDDRLVLFYSEAKMDIIKEEMETSMNIEIKSVQMK
DFSKETHPHLRNILTPLSLKSVNEKRQQKEQILPQIIIKGSIKPPPIQQSLFQISIDG
MFMIFVPHSWIPVQDSIMKLSKFATEAWNRYTIKVYGEAPTNDEIIQSGLSFLSLIK
DIKISIPGDCTVNSGQGENEYPLFIRSSLDIQTSRGAMGVEITLIDANNIQIYRNQ
LVDTNNENINNNNNNLTSTTTTTTNKKNQSNIKIIAPFKIVFQTTSTLLNSESSLS
INNIVVTFSYQDFKLMKIINSIIESNNQNLIQEKMKRNYLKKYRDPNECTSDDELEE
QLQQLENDDYNDDNYNSNGNNNNNNNSNNQLPVKEDSQLQKYLSISLEKGEFILNDDHK
ISSPIKLLSIGVDGLKSNIFSFPQKNQIALSLDANMKAGYFNKNIGIWEPLIENWGFS
FTSNNSIEGGWMVNFNSKIPLYINIKIFIDTSISTYQIWADDYSSQQKKDKKNKSND
YDNDEIIEDTVEQVNNDLIKPADDQKNNNNNNKNDNNNNNNNNNNNNNKNLQKEEIKYPYYIRNDT
NNNNNNNNNNNNNNNNNNIKNNNNNENNNNNKNLQKEEIKYPYYIRNDT
GVELWYWVRSDEIQKLPVGKEAGPNPKFSSSSSSSQQOQOPLSPLLQTQQQIQLPVK
ELQQKIELERKISFQLFGDFQPIAKVPMDSIGYTLYPMPEYKNIKLLYEISRNGAK
VLSLHSNFNIRNDTDIPVVVHITAIYNQIPRSIEBILLNPHVKIPIPVEYTIGRIKYK?
YNLGYDYSAERIDCSNVLQLFKSSKQKQLNNOLQQQSNNNNNNTNSNNNSNN
LLGNLIGGNNNNNDLNIGKISSKMICKHGVKLPFVFLSSIEKNSLETKNTIEISINTA
IMIENVLACDLQYRLYHGKNKKIIGSAFFSGVIPMGKKLPVLVYDSLQDIYMEFQIYD
FQWSPLCLIDSMVGLAVTDKVKIEDRLHQPLLISFDNRIQPNGSRFVTLYCEYWLINQ
TGLPLYFRHHIGAQTIDPAGQTPTKQIETIHESTSMKPTDSRNWYTKEWNHPSQPFMF
AYSDNSIVGGKFSLKIYDSNWSSPFSLISSQSSSNSNIEISEDRTDEEKEIKNVIGTI
TKRQPLKTNYNLSVSILPSNSKFWRTKVVTFSPMYLMVNSTRFRILYQOFDCDSNTQS
IIQDQSLPFQPPSSRHEKLIRIGILDTSAISNNPYNANDVSNIKWSGYFNPQILGQVV
LRLRNETDNSSGGVLNKFKKDSEISTKKRNSNFISNLLTIKNKQLNVDSSNQQQQLQ
QQQQQOQQOQQOQQOQQQSLQQQSIQQTSLNSSSSTITSTNSTVIFEDLRVFIN
VTIRVKSKSLTTMILLNEQNPELPPYRINNKTRFPLWIRQKKTEIWDKVQPKTSIP
YAWDHPILPKKLIEFPTGLTKTYRLGNLEENSIVSIKNPNSSGGYDRVDFQISIIAN
GPTRVLNIKERDLNGDTNNTISPQLNSIETFGTTNDPNPNPNNPNNNDKLNVKY
EFSIDLSNIGISIINKIPTREEIVYLSIDGLKLEVKQSKLDQYIQFKVDDLQIDDQRY
STNFPVFLCQSKKINHSINQQPNQQSSSPQSTNTTTTNTTTNNTTQQQKLKPFL
QFSATRTLKYSNIMFFRYFSILIQEFDINLDEASILNALSFININLNSLNEHFTLHPT
ITQEEILETKNASNIENHMVYFEMLHINPVKMNLSFISCKSPKETQAILGARSLAELL
IGFKSNSPFLNIERAPIKFNGFIWEHPLSTRQVIDEISLHFSYQMSQAHKIFGSFD
FIGNPIRLAESLGSGFKDFHEPALGLVKSPQDFAAGLSKGTSSLINNSVFGFADSTS
KITGTISKGLVQLSLDDSYIKEREQESNKQPKGVKEGLEFGFRDFGEGVIKGITGIID
EPYKGATQEKSWEGFFKGIGKGVLGVAVKPTVGVFDLVSKTSEGIKNSTTVAKSLSQI
KRRRIPRYFPREGTLSTVNQFKSIGSFLLYSKIGPPPTHDWYVFHTILNYKDLLLIGS
NYHLLIRCANLSINGNLSDSEIIWKVKFSEIIQIKNIPSNFGFQLFVNVGVGOPPI
TLIIPTLTEEVRNSVVRKITDLILLEKRFGNNLD"

BASE COUNT     5345 a    1560 c    1507 g    4256 t
ORIGIN

Query Match
Best Local Similarity   7.2%;   Score 44.2;   DB 3;   Length 12668;
Matches     85;   Conservative     0;   Mismatches     68;   Indels     1.8;   Gaps     0;

```
Qy    10 atgggatagaacagttcaacagtaagtacttcaaaaacaaagcatcatggcaagacag 69
         ||  || |||| ||  |||||  |||||| ||||||||||||||||||||||
Db  9280 ATTCATCAAATCAACAACAACAATTACACAACAACAACAACAACAACAACAAC 9339

Qy    70 atgggagtactgaataacaacaatggcgacaataacaacaataatgttacaagctca 129
         | |  |||||| |||||||||||||  ||||||||| ||||||  ||| ||||
Db  9340 AACAACAACAACAACAACAACAACAACAACAACAACAACAACAATCAATCAAC 9399

Qy   130 ttacgagtaattatggaaatgaagatcatta 162
         ||  | |  |||| || || |||||||
Db  9400 AAACTTCATTAAATTCATCATCATCAACAACTA 9432
```

us-09-156-580-1_copy_190_807.rge

OM nucleic - nucleic search, using sw model

Copyright   GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: July 15, 1999, 13:58:17 ; Search time 62 Seconds
                                      (without alignments)
                                      1875.243 Million cell updates/sec

Title: US-09-156-580-1_COPY_190_807
Perfect score: 618
Sequence: 1 atggcaggcatggatagaaa.........aacttcgactggatccact 618

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database :   N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result NO.	Score	% Query Match	Length	DB	ID	Description
	1	40.8	6.6	198	1	T17202	DNA-spanner oligon
	2	40.8	6.6	198	1	T17203	DNA-spanner oligon
	3	40.6	6.6	1690	1	Q92524	Nicotiana alata ar
	4	39.2	6.3	144	1	Q12515	CSP-2 peptide from
	5	39.2	6.3	198	1	T17204	DNA-spanner oligon
	6	39.2	6.3	198	1	T17205	DNA-spanner oligon
	7	37.6	6.1	120	1	Q12516	CSP-2 peptide from
	8	35.6	5.8	117	1	Q12514	CSP-2 peptide from
	9	35.4	5.7	9636	1	Q67190	P. falciparum tran
	10	35.4	5.7	1942	1	T38896	Carnation ACC synt
	11	35	5.7	2160	1	T85328	Truncated murine T
	12	34.8	5.6	1016	1	N93636	Sequence of alpha-
	13	33.8	5.5	8920	1	Q62924	Carbamoyl-phosphat
c	14	33.8	5.5	58407	1	V21210	Methanococcus jann
c	15	33.6	5.4	780	1	V26001	Bacteriophage DNA
c	16	33.6	5.4	2519	1	V54124	Human membrane pro
c	17	33.4	5.4	110000	1	T58840_2	Continuation (3 of
c	18	33.4	5.4	110000	1	T58840_3	Continuation (4 of
	19	33.2	5.4	3600	1	T77330	Solanum tuberosum
	20	33.2	5.4	1813	1	V34193	Human secreted pro
	21	33	5.3	9295	1	T62125	Arabidopsis thalia
	22	33	5.3	5318	1	V20701	Cryptosporidium pa
	23	33	5.3	5163	1	V20700	Cryptosporidium pa
	24	32.8	5.3	3101	1	Q02047	Sequence encoding
	25	32.8	5.3	1430	1	Q92525	Nicotiana plumbagi
	26	32.8	5.3	1040	1	Q92530	P. communis (pear)
	27	32.6	5.2	2183	1	Q54769	T. parva sporozoit
c	28	32.6	5.2	15229	1	V18276	RSV isolate 18537
	29	32.4	5.2	4467	1	T68648	pSRQ800 fragment 1
	30	32.4	5.2	15225	1	V17552	Respiratory syncyt
c	31	32.4	5.2	15219	1	V18277	RSV vaccine 2B33F
c	32	32.4	5.2	15219	1	V18278	RSV vaccine 2B20L
c	33	32.4	5.2	15219	1	V18279	RSV revertant 2B33
c	34	32.4	5.2	15219	1	V18280	RSV revertant 2B20
c	35	32.4	5.2	15218	1	V18275	RSV isolate 2B wil
c	36	32.4	5.2	110000	1	V21209_15	Continuation (16 o
	37	32.2	5.2	11065	1	Q94335	Degenerate Alterom
	38	32	5.2	7241	1	Q15140	Genomic clone LE-A
c	39	32	5.2	6106	1	Q74678	Bacillus cereus ve
c	40	32	5.2	6049	1	T13939	B. cereus VIP2A(a)
c	41	32	5.2	2417	1	T51537	Mycobacterium gall
c	42	32	5.2	350	1	T65080	Canine genomic mic
c	43	32	5.2	6049	1	T73993	B. cereus VIP2A(a)
c	44	32	5.2	6049	1	V16165	DNA encoding veget
c	45	32	5.2	7244	1	V15705	Tomato ACC synthas

ALIGNMENTS

RESULT   1
T17202
ID   T17202 standard; DNA; 198 BP.
AC   T17202;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide BAM36-198.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;
KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9605847-A1.
PD   29-FEB-1996.
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3; Page 18; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-1980p, RI36-198
CC   and RI36-1980p, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-1980p and
CC   RI36-1980p, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP;   121 A;   64 C;   7 G;   6 T;

Query Match   6.6%; Score 40.8; DB 1; Length 198;
Best Local Similarity   61.1%; Pred. No. 0.038;
Matches   66; Conservative   0; Mismatches 42; Indels   0; Gaps   0;

Qy   9 catggtgatgaaacagtttcacagtagtacttcaaaacaaaagcatcatggcaagaca 68
              || |||| |||| |||| ||| |||||| |||| ||||| | |||||||| |||||
Db   34 CAAGGATCCCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 93

Qy   69 gatggagtacttgaataacaacaatggcgacaataacaacaacaataa 116
              || ||| || || |||||||||||||| |||||||||||||||||||
Db   94 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 141

RESULT   2
T17203
ID   T17203 standard; DNA; 198 BP.
AC   T17203;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide BAM36-1980p.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;

KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9505847-A1.
PD   29-FEB-1996.
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI: 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3; Page 18; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM-36-198op, RI36-198
CC   and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC   RI36-198op, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP;   121 A;   64 C;   7 G;   6 T;

Query Match   6.6%;   Score 40.8;   DB 1;   Length 198;
Best Local Similarity   61.1%;   Pred. No. 0.038;
Matches   66; Conservative   0; Mismatches   42; Indels   0; Gaps   0;

Qy   9 catggatagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagaca 68
         || |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   34 CAAGGATCCCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 93

Qy   69 gatggagtacttgaataacaacaatggcgcacaataacaacaacaataacaacaacaataa 116
         | | || |||||||||| ||||||| ||||||||||||| ||
Db   94 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 141

RESULT 3
Q92524
ID   Q92524 standard; DNA; 1690 BP.
AC   Q92524;
DT   05-DEC-1995 (first entry)
DE   Nicotiana alata arabinogalactan protein cDNA NaAGP1.
KW   Arabinogalactan protein; AGP; gum; thickening agent; emulsifier;
KW   adhesive; ss.
OS   Nicotiana alata.
FH   Key   Location/Qualifiers
FT   cds   60. .1445
FT   /*tag= a
FT   misc_difference 1. .31
FT   /*tag= b
FT   /note= "PCR SEQ non-overlapping with cDNA SEQ"
PN   WO9515377-A.
PD   08-JUN-1995.
PF   01-DEC-1994; AU0744.
PR   01-DEC-1993; US-161944.
PR   18-JUL-1994; US-276452.
PA   (ALBR) ALBRIGHT & WILSON AUSTRALIA.
PI   Bacic A, Chen C,   Clarke AE, Du H,   Gane AM,   Mau S;
DR   WPI: 95-215258/28.
DR   P-PSDB; R75506.
PT   Recombinant plant arabinogalactan protein and cloned DNA encoding it -

PT   useful as an emulsifying agent, adhesive agent or a lubricating agent.
PS   Claim 8; Figure 1F; 142pp; English.
CC   The N-terminal region of an arabinogalactan protein (AGP) was obtd.
CC   from N. alata suspension culture. It was used as a template for
CC   the synthesis of an oligo primer which was further utilised for the
CC   isolation of a hybridising AGP gene from both N. alata and N.
CC   plumbaginafolia. Degenerate reverse primers corresp. to part of the
CC   AGP N-terminal AA sequence were synthesised and used to obtain a
CC   160 bp primer extension product, which was then amplified by PCR.
CC   Additional primers, corresp. in sequence to parts of the 160-bp
CC   fragment were synthesised and used to amplify the 3' part of the
CC   AGP gene by nested PCR. The primers were Q92535 and Q92537. A
CC   1.6 kb fragment was amplified and sequenced. The alignment of the
CC   sequences obtd. from the two PCR reactions gave rise to a DNA
CC   sequence shown. The PCR fragment encodes a protein contg.
CC   the peptide (R75513) by protein sequencing with two mismatches
CC   (see Q92524 FT). The 1.6 kb PCR fragment was used to screen a cDNA
CC   library made from RNA isolated from N. alata cells. The alignment
CC   of the PCR sequences with the cDNA sequence gave rise to Q92524.
CC   The calc. mol. wt. of the encoded protein is 51.8 kD with a
CC   predicted pI of 3.84.
SQ   Sequence 1690 BP;   619 A;   352 C;   274 G;   445 T;

Query Match   6.6%;   Score 40.6;   DB 1;   Length 1690;
Best Local Similarity   58.8%;   Pred. No. 0.081;
Matches   70; Conservative   0; Mismatches   49; Indels   0; Gaps   0;

Qy   9 catggatagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagaca 68
         || |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db 1022 CAATAACAATAAATGTTTCTCCGAGAATTACAACAACAATAACAACAATGCTTTTCTA 1081

Qy   69 gatggagtacttgaataacaacaatggcgcacaataacaacaacaataatgttaccaagct 127
         | | ||| ||| ||||||||||||| ||| ||||||||||||||| |||| || ||
Db 1082 CGAGAATTACAACAACAACAACAATGGCTACTCAGAGAACTACAACTCAGGCTAGCAGCT 1140

RESULT 4
Q12515
ID   Q12515 standard; DNA; 144 BP.
AC   Q12515;
DT   18-SEP-1991 (first entry)
DE   CSP-2 peptide from P. falciparum (clone 4).
KW   P. berghei; Pf42; Pf54; Pb42; Pb54; ss.
OS   Plasmodium falciparum.
PN   WO9108756-A.
PD   27-JUN-1991.
PF   12-DEC-1990; U07194.
PR   12-DEC-1989; US-448109.
PA   (BIOM-) BIOMEDICAL RES INST.
PI   Hollingdal MR, Sina B;
DR   WPI: 91-207865/28.
DR   P-PSDB; R12903.
PT   Sporozoite polypeptide antigens - are useful for antimalarial
PT   vaccines and the encoding these peptide(s) can be incorporated
PT   into virus to produce live vaccines
PS   Disclosure; Page 21; 32pp; English.
CC   A portion of the gene encoding the P. falciparum sporozoite
CC   antigenic protein CSP-2 (mol. wt. 42kD/54kD) was cloned. Clones 3,
CC   4 and 6 were identified and found to contain malaria DNA inserts as
CC   represented in Q12514, Q12515 and Q12516 respectively. The
CC   polypeptides are present on the surface of the sporozoites and are
CC   cross-reactive with P. berghei antiserum. The Pf and Pb
CC   proteins are substantially homologous, if not identical.
SQ   Sequence 144 BP;   87 A;   48 C;   9 G;   0 U;

Query Match   6.3%;   Score 39.2;   DB 1;   Length 144;
Best Local Similarity   60.2%;   Pred. No. 0.09;
Matches   65; Conservative   0; Mismatches   43; Indels   0; Gaps   0;

Qy   9 catggatagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagaca 68
         || |||| |||| |||| ||||| || |||||||||||| |||||||| |

Db   30 CAGCAACAGCCAACAACAACAACAACAACAACAACAGCAACAACAACAA 89

Qy   69 gatggagtacttgaataacaacaacatggcgacaataacaacaacaataa 116
          || || ||||||| || ||| |||||||| |||||||||||||| |
Db   90 CAACAACAACAACAGCCAACAACAACAGCAACAACAACAACAACAACAACAA 137

RESULT 5
T17204
ID   T17204 standard; DNA; 198 BP.
AC   T17204;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide RI36-198.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid;
KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic
PN   WO9605847-A1.
PD   29-FEB-1996.
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3; Page 19; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-198op, RI36-198
CC   and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence,
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC   RI36-198op, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP; 122 A; 63 C; 6 G; 7 T;

Query Match   6.3%; Score 39.2; DB 1; Length 198;
Best Local Similarity 60.2%; Pred. No.0.099;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 9 catggagtagaaacagtttcaacagtagtacttcaaaaacaaaagcatcatggccaagaca 68
       || | || ||||| |||||| ||| || ||||||| | |||||| ||| ||
Db 10 CAACAACAACAACAACAACAACAACAACAGAATTCCAACAACAACAACAACAACAACAACA 69

Qy 69 gatggagtacttgaataacaacaacatggcgacaataacaacaacaataa 116
       | | ||| || ||||||| | |||| ||||||||||| ||
Db 70 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACACA 117

RESULT 6
T17205
ID   T17205 standard; DNA; 198 BP.
AC   T17205;
DT   03-OCT-1996 (first entry)
DE   DNA-spanner oligonucleotide RI36-198op.
KW   DNA-spanner oligonucleotide; restriction site; anneal; hybridise;
KW   complementary sequence; anti-test protein antibody; extension;
KW   primer template complex; double stranded duplex; amino acid; .

KW   sequence determination; epitope ordering; restriction mapping; ss.
OS   Synthetic.
PN   WO9605847-A1.
PD   29-FEB-1996.
PF   22-AUG-1995; U10668.
PR   22-AUG-1994; US-294133.
PA   (UYPE-) UNIV PENNSYLVANIA.
PI   Eberwine J;
DR   WPI; 96-151138/15.
PT   Sequencing proteins by epitope ordering and protein restriction
PT   mapping - allows for characterisation of post-translational events
PT   on the protein
PS   Example 3; Page 19; 28pp; English.
CC   The DNA-spanner oligonucleotides BAM-36-198, BAM36-198op, RI36-198
CC   and RI36-198op, comprise a modified 5'-nucleotide followed by a
CC   restriction site, and a long nucleotide sequence contg. a sequence
CC   at the 3'-end capable of annealing to a complementary sequence.
CC   i.e. BAM36-198 and RI36-198 are complementary to BAM36-198op and
CC   RI36-198op, respectively. Binding a complementary pair of oligos to
CC   an anti-test protein antibody (Ab), which is then bound to the
CC   protein, allows the oligo sequences to form a primer template
CC   complex, which can be extended to form a double stranded duplex.
CC   The duplex can be digested with a restriction enzyme, the prods.
CC   separated and their band pattern detected, allowing a portion of
CC   the protein's amino acid sequence to be determined.
CC   This method allows the determination of a protein sequence by
CC   epitope ordering with single, di- and tri-amino acid specific Abs,
CC   followed by restriction mapping. The procedure provides enough
CC   protein sequencing to allow library screening oligo probes to be
CC   generated, and the corresp. cDNA clone to be PCR isolated. In addn.
CC   the epitope ordering procedure allows post-translational
CC   modification events to be characterised.
SQ   Sequence 198 BP; 122 A; 63 C; 6 G; 7 T;

Query Match   6.3%; Score 39.2; DB 1; Length 198;
Best Local Similarity 60.2%; Pred. No. 0.099;
Matches 65; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 9 catggagtagaaacagtttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 68
       || | || ||||| |||||| ||| || ||||||| | |||||| ||| ||
Db 10 CAACAACAACAACAACAACAACAACAACAGAATTCCAACAACAACAACAACAACAACAACAA 69

Qy 69 gatggagtacttgaataacaacaacatggcgacaataacaacaacaataa 116
       | | ||| || ||||||| | |||| ||||||||||| ||
Db 70 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 117

RESULT 7
Q12516
ID   Q12516 standard; DNA; 120 BP.
AC   Q12516;
DT   18-SEP-1991 (first entry)
DE   CSP-2 peptide from P. falciparum (clone 6).
KW   P. berghei; Pf42; Pf54; Pb42; Pb54; ss.
OS   Plasmodium falciparum.
PN   WO9108756-A.
PD   27-JUN-1991.
PF   12-DEC-1990; U07194.
PR   12-DEC-1989; US-448109.
PA   (BIOM-) BIOMEDICAL RES INST.
PI   Hollingdal MR, Sina B;
DR   WPI; 91-207865/28.
DR   P-PSDB; R12904.
PT   Sporozoite polypeptide antigens - are useful for antimalarial
PT   vaccines and the encoding these peptide(s) can be incorporated
PT   into virus to produce live vaccines
PS   Disclosure; Page 22; 32pp; English.
CC   A portion of the gene encoding the P. falciparum sporozoite
CC   antigenic protein CSP-2 (mol. wt. 42kD/54kD) was cloned. Clones 3,
CC   4 and 6 were identified and found to contain malaria DNA inserts as
CC   represented in Q12514, Q12515 and Q12516 respectively. The
CC   polypeptides are present on the surface of the sporozoites and are

CC   cross-reactive with P. berghei antiserum. The Pf and Pb
CC   proteins are substantially homologous, if not identical.
SQ   Sequence 120 BP;   40 A;   70 A;   10 G;   0 U;

Query Match   6.1%; Score 37.6; DB 1; Length 120;
Best Local Similarity 59.3%; Pred. No. 0.22;
Matches   64; Conservative   0; Mismatches   44; Indels   0; Gaps   0;

Qy   19 aacagtttcaacagtaagtacttcaaaacaaaaagcatcatggcaagacagatggagtac 78
         | ||| |||||| ||| || |||||| |||| || | |||
Db   1 AGCAGCCAACAACAACACAGCAGCAACACAGCAACCAACAGCAACAACAACAACAACAAC 60

Qy   79 ttgaataacaacaatggcgacaataacaacaacaataatgttacaagc 126
         || ||||||| || ||||||||| || |||
Db   61 AACAACAACAACAACAGCAGCAACACAGCAGCAACAACAACAACAACAGC 108

RESULT 8
Q12514
ID   Q12514 standard; DNA; 117 BP.
AC   Q12514;
DT   18-SEP-1991 (first entry)
DE   CSP-2 peptide from P. falciparum (clone 3).
KW   P. berghei; Pf42; Pf54; Pb42; Pb54; ss.
OS   Plasmodium falciparum.
PN   WO9108756-A.
PD   27-JUN-1991.
PF   12-DEC-1990; U07194.
PR   12-DEC-1989; US-448109.
PA   (BIOM-) BIOMEDICAL RES INST.
PI   Hollingdal MR, Sina B;
DR   WPI; 91-207865/28.
DR   P-PSDB; R12902.
PT   Sporozoite polypeptide antigens - are useful for antimalarial
PT   vaccines and the encoding these peptide(s) can be incorporated
PT   into virus to produce live vaccines
PS   Disclosure; Page 21; 32pp; English.
CC   A portion of the gene encoding the P. falciparum sporozoite
CC   antigenic protein CSP-2 (mol. wt. 42kD/54kD) was cloned. Clones 3,
CC   4 and 6 were identified and found to contain malaria DNA inserts as
CC   represented in Q12514, Q12515 and Q12516 respectively. The
CC   polypeptides are present on the surface of the sporozoites and are
CC   cross-reactive with P. berghei antiserum. The Pf and Pb
CC   proteins are substantially homologous, if not identical.
SQ   Sequence 117 BP;   69 A;   39 C;   9 G;   0 U;

Query Match   5.8%; Score 35.6; DB 1; Length 117;
Best Local Similarity 60.2%; Pred. No. 0.72;
Matches   59; Conservative   0; Mismatches   39; Indels   0; Gaps   0;

Qy   19 aacagtttcaacagtaagtacttcaaaacaaaaagcatcatggcaagacagatggagtac 78
         | ||| |||||| ||| || |||||| |||| || | |||
Db   1 AGCAGCCAACAACAACACAGCAGCAACACAGCAGCAACAACAACAACAACAACAACAAC 60

Qy   79 ttgaataacaacaatggcgacaataacaacaacaataa 116
         || ||||||| || |||||||| |||||||||||
Db   61 AACAACAACAACAACAGCAGCAACACAGCAGCAACAACAACAA 98

RESULT 9
Q67190
ID   Q67190 standard; DNA; 9636 BP.
AC   Q67190; AU0286.
DT   20-FEB-1995 (first entry)
DE   P. falciparum transmission blocking target antigen Pfs230 DNA.
KW   Protozoon; transmission blocking target antigen; Pfs230; malaria;
KW   vaccine; ss.
OS   Plasmodium falciparum.
FH   Key   Location/Qualifiers
FT   cds   149..9553
FT   /tag= a

PN   WO9417187-A.
PD   04-AUG-1994.
PF   18-JAN-1994; U00547.
PR   29-JAN-1993; US-010409.
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA   (USSH) US SEC DEPT HEALTH.
PI   Kaslow DC, Williamson KC;
DR   WPI; 94-264101/32.
DR   P-PSDB; R57474.
PT   New Plasmodium falciparum transmission blocking target antigen -
PT   useful in antimalarial vaccines, also related DNA, expression
PT   vectors and transformed cells
PS   Claim 2; Page 24; 63pp; English.
CC   The DNA may be used to express Psf230 in a host cell. It can also
CC   be used in vaccines (by incorporation into viral vectors which are
CC   then used to infect host cells) and oligonucleotides derived from it
CC   can be used to identify homologous proteins in other spp.
SQ   Sequence 9636 BP;   4156 A;   974 C;   1422 G;   3084 T;

Query Match   5.7%; Score 35.4; DB 1; Length 9636;
Best Local Similarity 46.8%; Pred. No.3;
Matches 111; Conservative   0; Mismatches 126; Indels   0; Gaps   0;

Qy   82 aataacaacaatggcgacaataacaacaacaataatgttacaagctcattacgagataat 141
         ||| |||||| || ||| |||| || || |||||
Db 8177 AATCAACAACAATGATGCTATTAGTAATAATAATAATAATTCATATTCTATATTACACAT 8236

Qy   142 tatggaaatgaagatcattacttggtggactatctcttggcctccaagatcttataca 201
         | ||| |||| || ||| || |||||
Db 8237 AATAAAAATACAGAGAATAATCTAATAATGTGATATATCTTTAATTCCAAAAACTGTTATA 8296

Qy   202 tgtagctttgtaaaagggattagatctgctgctcaagcctcttggtggacacatgaatgtt 261
         |||||| ||| |||| ||| || | |||
Db 8297 GGAATTAAATGTCCTAATAAAAAATTAAATCCACAAACATGTTTTGATGAAGTGTATTAT 8356

Qy   262 catagaagagatagagcccatttggacaatcaccaccactagagatattaataggtat 318
         ||| || |||
Db 8357 GTTAAACAAGAAGATGTACCTTCGAAAACTATAACAGCTGATAAATATATACACATTT 8413

RESULT 10
T38896
ID   T38896 standard; DNA; 1942 BP.
AC   T38896;
DT   10-FEB-1997 (first entry)
DE   Carnation ACC synthase gene.
KW   ACC synthase; 1-aminocyclopentane-1-carboxylic acid synthase;
KW   ACC synthase; co-suppression; ethylene; senescence;
KW   transgenic plant; ss.
OS   Dianthus sp. cv. White Sim.
FH   Key   Location/Qualifiers
FT   cds   134..1687
FT   /tag= a
FT   primer_bind complement (287..308)
FT   /tag= b
FT   /note= "5' primer for truncated ACC oxidase gene"
FT   primer_bind 1352..1373
FT   /tag= c
FT   /note= "3' primer for truncated ACC oxidase gene"
PN   WO9635792-A1.
PD   14-NOV-1996.
PF   09-MAY-1996; AU0286.
PR   09-MAY-1995; AU-002862.
PA   (ALLR-) ALLRAD NO 1 PTY LTD.
PA   (FLOR-) FLORIGENE INVESTMENTS PTY LTD.
PI   Cornish EC, Graham MW, Gutterson NI, Michael MZ;
PI   Tucker WT;
DR   WPI; 96-518680/51.
DR   P-PSDB; W04558.
PT   Producing transgenic plants, with reduced climacteric ethylene
PT   prodn. - to give flowers and buds, specifically carnations, with
PT   increased post-harvest life

PS Claim 9: Page 39-42; 98pp; English.
CC A full-length gene (T38896) codes for carnation cv White Sim
CC 1-aminocyclopentane-1-carboxylic acid (ACC) synthase (W04558),
CC an enzyme involved in the biosynthesis of ethylene. Expression
CC of an internal, non-full-length fragment (T38897) of the gene
CC in transgenic carnation reduces the formation of ACC synthase
CC by co-suppression. The reduction in ACC synthase activity results
CC in decreased climacteric ethylene prodn. by the plant and hence
CC delayed senescence of flowers and buds after cutting, and increased
CC post-harvest life.
SQ Sequence 1942 BP;　625 A;　371 C;　383 G;　563 T;

Query Match　　　　　5.7%;　Score 35.4;　DB 1;　Length 1942;
Best Local Similarity　61.3%;　Pred. No. 1.9;
Matches　57;　Conservative　0;　Mismatches　36;　Indels　0;　Gaps　0;

Qy　　25 ttcaacagtagtacttcaaaaacaaaagcatcggcaagacagatggagtacttgaat 84
          |||||||| || |||| ||||||||||||||||||||||||||||||||||||||||
Db　1453 TTCAACAATGACACAACATCAGCAAGACGCAGCAACAACAACAACAACAACAAC 1512

Qy　　85 aacaacaatggcgacaataacaacaacaatat 117
          ||||||||| |||| ||||||||||||| |||
Db　1513 AACAACAACAACAACAACAACAACAACAACGACAAT 1545

RESULT 11
T85328
ID T85328 standard; cDNA; 2160 BP.
AC T85328;
DT 08-DEC-1997 (first entry)
DE Truncated murine TRAF inhibitor protein I-TRAFalpha encoding cDNA.
KW Tumour necrosis factor receptor associated factor; TNF; tumour;
KW immunosuppressant; antiinflammatory; Epstein Barr virus;
KW septic shock; rheumatoid arthritis; Burkitt's lymphoma;
OS Mus musculus.
OS nasopharyngeal carcinoma; human; mouse; ss.
PN WO9707134-A1.
PD 27-FEB-1997.
PF 14-AUG-1996; U13129.
PR 17-AUG-1995; US-002382.
PA (GETH) GENENTECH INC.
PA (TULA-) TULARIK INC.
PI Goeddel DV, Rothe M;
DR WPI; 97-165244/15.
PT Inhibitor of tumour necrosis factor associated factor - useful to
PT develop immunosuppressants, antiinflammatory agents and agents for
PT treating Epstein Barr virus associated tumours
PS Disclosure; Page 32-33; 43pp; English.
CC The present sequence encodes a novel isolated truncated murine
CC inhibitor of tumour necrosis factor associated factor (I-TRAF)
CC polypeptide, I-TRAF. The protein is capable of inhibiting a TRAF-
CC mediated signalling event. A tumour associated with Epstein-Barr virus
CC (EBV) can be treated by administering an I-TRAF to a patient having
CC developed or at risk of developing such tumour. The I-TRAF polypeptide,
CC which inhibits the interaction of TRAF with members of the TNF receptor
CC superfamily, can be used to develop immunosuppressants, antiinflammatory
CC agents, or agents for the treatment of septic shock, rheumatoid
CC arthritis or EBV-associated tumours, e.g. Burkitt's lymphoma and
CC nasopharyngeal carcinoma.
SQ Sequence 2160 BP;　720 A;　426 C;　404 G;　610 T;

Query Match　　　　　5.7%;　Score 35;　DB 1;　Length 2160;
Best Local Similarity　59.8%;　Pred. No. 2.4;
Matches　76;　Conservative　0;　Mismatches　50;　Indels　1;　Gaps　1;

Qy　　3 ggcaggcatggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggc 62
         | |||||| || ||||| || |||| || |||| ||||| || || ||||| || |
Db　142 GACAGGCATGCATGGATAGAGATTCAGCAGTAAG-AGAGCTACAGCAAAAGACTGAGAAC 200

Qy　 63 aagacagatggagtacttgaataacaacaatggcgacaataacaacaacaatgttac 122
         || |||||| ||||||| || || || |||| || |||| || ||||||| |

Db　201 TATGACAAAGAATACGGAGCAAGGACAGGAACAGCAGCTGTGTCATTTCAACAAAACCTAATTGAC 260

Qy　123 aagctca 129
         | ||||
Db　261 AGGCTGA 267

RESULT 12
N93636
ID N93636 standard; DNA; 1016 BP.
AC N93636;
DT 22-MAR-1991 (first entry)
DE Sequence of alpha-gliadin gene fragment used to construct plasmid
DE pAY32
KW Coeliac disease diagnosis; dough formation; food technology; ds.
OS wheat.
PN US482765-A.
PD 02-MAY-1989.
PF 24-OCT-1986; 922616.
PR 24-OCT-1986; US-922616.
PA (UYHA-) Univ of Hawaii (USDA).
PI Greene FC, Stiles JI, Neill JD, Anderson OD, Litts JC;
DR WPI; 89-150270/20.
PT New wheat gluten protein gene plasmid - useful for genetic
PT transformation of yeasts to produce gliadin or glutenin
PS Example; Fig 5; 21pp; English.
CC The preferred gene fragment in the patent of the invention encodes
CC alpha-gliadin or glutenin, and the preferred plasmid encoding alpha-
CC gliadin is PAY 31. The plasmid may be used for the prodn. of alpha-
CC gliadin, which may be used in the food technology industry to modify
CC prodn. processes of doughs and batters and to reduce unit costs.
CC Gliadin may also be used for diagnosis and treatment of illness
CC caused by wheat gluten proteins and for testing theories of dough
CC formation.
SQ Sequence 1016 BP;　359 A;　321 C;　141 G;　195 T;

Query Match　　　　　5.6%;　Score 34.8;　DB 1;　Length 1016;
Best Local Similarity　52.8%;　Pred. No. 2.2;
Matches　75;　Conservative　0;　Mismatches　67;　Indels　0;　Gaps　0;

Qy　 41 tcaaaaacaaaagcatcatggcaagacagatggagtacttgaataacaacaatggcgaca 100
         || |||||||| || || || |||||| |||||| ||| || || ||||||| |||
Db　345 TCTACACCACCACACAACCAATTCGCAGCAACAAGCAACAACAACAACAACAACAACACA 404

Qy　101 ataacaacaacaataatgttacaagctcattacgagataataattatgtgaaatgaagatcatt 160
         | ||||||||||| || || | || ||| | |||| ||||| ||| | ||| ||| |||
Db　405 ACAACAACAACAACAACAACAACAACAATCCTTCAACAAATTTTGCAACAACAACTGAT 464

Qy　161 tacttggtggactattcctctg 182
         || ||| || || ||||| ||
Db　465 TCCATGCAGGGATGTGTCTTG 486

RESULT 13
Q62924
ID Q62924 standard; cDNA; 8920 BP.
AC Q62924;
DT 06-DEC-1994 (first entry)
DE Carbamoyl-phosphate-synthetase II.
KW Carbamoyl-phosphate-synthetase II; CPSII; pscPSII gene;
KW malaria; ss.
OS Plasmodium falciparum.
FH Key   Location/Qualifiers
FT cds   1226..8401
FT   /*tag= a
FT   /EC_number= 6.3.5.5
PN WO9412643-A.
PD 09-JUN-1994.
PF 02-DEC-1993; AU0617.
PR 02-DEC-1993; AU-006206.
PR 16-DEC-1992; AU-006380.
PA (UNIX) UNISEARCH LTD.

PI   Flores MV, Osullivan WJ, Stewart TS;
DR   WPI: 94-200271/24.
DR   P-PSDB: R55694.
PT   Nucleic acid encoding carbamoyl phosphate synthetase II -
PT   isolated from Plasmodium falciparum, used to develop prods. for
PT   the treatment of malaria.
PS   Disclosure: Page 6-16; 31pp; English.
CC   The cDNA sequence encoding the carbamoyl-phosphate-transferase II
CC   (CPSII) of Plasmodium falciparum was determined. The cDNA encodes
CC   a protein that includes 2 insert sequences not found in other CPSII
CC   proteins. The first separates the putative structural subdomain and
CC   the glutaminase subdomain of the glutamine-amidotransferase subunit
CC   of CPSII, while the second separates 2 ATP binding subdomains of the
CC   CPSII subunit, CPSa and CPSb.
SQ   Sequence 8920 BP; 3836 A; 774 C; 1232 G; 3078 T;

Query Match   5.5%; Score 33.8; DB 1; Length 8920;
Best Local Similarity 48.7%; Pred No. 7.6;
Matches 92; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy   28 aacagtaagtacttcaaaaacaaaagcatcatggcgaagacagagatggagtacttgaatac 87
         || | || ||| ||||||||||||||| |||||||||||||||||||||
Db 655 AAAATTTACATATGAAAAATGAACTTGTATATGTAAATTTATAAATATTTAAACATAAA 714

Qy   88 aacaatggcgacaataacaacaacaataatgttacaagctcattacgagataattatga 147
         || ||||||| |||| |||||||| || ||||||| ||| |||
Db 715 TATAAAATGTATAAAAAAAAAAAAAAAAGAAAAATGGGAAAAAATATATATAGATTATATATAATAAA 774

Qy 148 aatgaagatcattacttggtggactattcctcttggcctccaagatcttatacatgtagc 207
          |||||| || |||||||||| || || |||||||| ||||
Db 775 TATATATATATATAATATTGGGGATATTCCTGAATCATAGGTCTTAAACAGTTTA 834

Qy 208 tttgtaa 216
         |||||||
Db 835 TTCTTTTAA 843

RESULT   14
V21210/c
ID   V21210 standard; DNA; 58407 BP.
AC   V21210;
DT   10-NOV-1998 (first entry)
DE   Methanococcus jannaschii large circular extrachromosomal element.
KW   Methanococcus jannaschii; methanogenic archaeon; circular chromosome:
KW   genome; autotrophic; extrachromosomal element; identification; ds.
OS   Methanococcus jannaschii.
PN   WO9807830-A2.
PD   26-FEB-1998.
PF   22-AUG-1997; U14900.
PR   22-AUG-1996; US-024428.
PA   (GENO-) INST GENOMIC RES.
PA   (UNI-) UNIV ILLINOIS FOUND.
PA   (UJJO-) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PI   Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
DR   WPI: 98-169145/15.
PT   Complete genome sequence of methano-genic archaeon, Methanococcus
PT   jannaschii - useful in identification of M. jannaschii genome
PT   fragment.
PS   Claim 13; Page 585-600; 614pp; English.
CC   The present sequence represents the large circular extrachromosomal
CC   element sequence of the Methanococcus jannaschii circular chromosome. The
CC   present invention describes M. jannaschii open reading frames from the
CC   genome sequence. The invention also describes a computer based system
CC   for identifying fragments of the M. jannaschii genome that are
CC   homologous to target nucleotide sequences, comprising: (a) data storage
CC   means comprising the nucleotide sequence of the 164976, 58407 or 16550
CC   bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
CC   least 99.9% identical to it; (b) search means for comparing a target
CC   sequence to the nucleotide sequence of the data storage means to
CC   identify a homologous sequence, and (c) retrieval means for obtaining
CC   the homologous sequence. The method, which is based on whole genome
CC   random sequencing of an autotrophic archaeon M. jannaschii, the genome

CC   of which consists of 3 physically distinct elements, a large circular
CC   chromosome (the 1664976 bp sequence given in V21209), a large circular
CC   extra-chromosomal element (the 58407 bp sequence given in V21210), and a
CC   small circular extra-chromosomal element (the 16550 bp sequence given in
CC   V21211), can be used in the identification of M. jannaschii genome
CC   fragment.
SQ   Sequence 58407 BP; 20163 A; 9034 C; 7447 G; 21761 T;

Query Match   5.5%; Score 33.8; DB 1; Length 58407;
Best Local Similarity 49.7%; Pred. No. 13;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy   33 taggtacttcaaaaacaaaagcatcatggcgaagacagatggagtacttgaataacaacaa 92
         ||| ||||||||||||||||| |||| ||||||||||||||
Db 23509 TATGTAAATACACAATACACAGAACAGAAGCATGTAAAGACTTAATTAAAAGATTAAGCAAAGCTA 23450

Qy   93 tggcgacaataacaacaacaataatgttacaagctcattacgagataattatgaaaatga 152
         ||| | || ||| |||| |||||||||
Db 23449 TCAACAACACGAACATAGATACATTGGAAATATCAATTGAAATCAATATTAACGGTA 23390

Qy 153 agatcattacttggtggactattcctcttggcctcccaagatcttatacatgta 205
         |||| |||||||| || |||||
Db 23389 AAATCATTATCGTTAGTGAAGATGAAGTTATCGACCTCATCCTTAAAGGTGAA 23337

RESULT   15
V26001
ID   V26001 standard; DNA; 780 BP.
AC   V26001;
DT   28-AUG-1998 (first entry)
DE   Bacteriophage DNA SEQ ID NO:1-2.
KW   Bacteriophage; pathogen; Escherichia coli 0157; bactericide; food;
KW   sterilisation; enteric haemorrhage; ds.
OS   Bacteriophage.
PN   WO9808944-A1.
PD   05-MAR-1998.
PF   26-AUG-1997; J02957.
PR   19-APR-1997; JP-135716.
PR   26-AUG-1996; JP-261132.
PR   14-APR-1997; JP-130236.
PA   (BIOV-) BIO VENTURE BANK CO LTD.
PI   Nishikori K, Takahashi S;
DR   WPI; 98-230262/20.
PT   Bacteriophage with high specificity for particular pathogens such as
PT   Escherichia coli 0157 - is incorporated in bactericides for food
PT   sterilisation
PS   Claim 3; Page 31; 54pp; Japanese.
CC   The present sequence represents a DNA sequence from a novel
CC   bacteriophage which has high specificity for pathogenic bacteria,
CC   especially for strains of Escherichia coli causing enteric haemorrhage,
CC   such as E. coli 0157. The bacteriophage can be incorporated into
CC   bio-bactericidal compositions. These can contain more than one
CC   bacteriophage strain in order to have a bactericidal effect against more
CC   than one pathogen at the same time. The bio-bactericide may also contain
CC   an amino-acid or other material to control the pH to 6.5-7.5 and ensure
CC   stability of the bacteriophage, such as glycine, arginine or lysine.
CC   The bio-bactericidal composition can be used for treating food (such as
CC   meat or fish) to prevent bacterial contamination, e.g. in fresh food or
CC   in the kitchens of restaurants, schools and other institutions. It may
CC   be formulated as a spray. It can also be used to sterilise working
CC   surfaces, aprons and other clothing. The bio-bactericides are safe to
CC   humans but very potent against pathogenic bacteria.
SQ   Sequence 780 BP; 261 A; 125 C; 151 G; 242 T;

Query Match   5.4%; Score 33.6; DB 1; Length 780;
Best Local Similarity 59.4%; Pred. No. 4.2;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 213 taaaggaatttagatctgctcaagctctgggacacatgaatgttcatagaagaga 272
         || | |||||||| | || ||||||||| || ||||| || |||
Db 186 TATTATTGAATTAGTAATTTCATATCCTCACGATGCCACATGAGAATATTGATACTATTGA 245

Fri Jul 16 12:11:59 1999

us-09-156-580-1_copy_190_807.rng

```
Qy   273 tagagccatttgagacaatcaccacctagagatat 308
         ||  |-||||| |-| |||||| | |-| |||||
Db   246 TAACTACACTTGCGTTATAAACCACTAAAAGATAT 281
```

Search completed: July 15, 1999, 13:58:23
Job time: 3939 sec

Copyright   GenCore version 4.5
            Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 1999, 12:07:21 ; Search time 377.02 Seconds
                               (without alignments)
                               2500.430 Million cell updates/sec

Title: US-09-156-580-1_COPY_190_807
Perfect score: 618
Sequence: 1 atggcaggcatggatagaaa........aacttcgactggatccact 618

Scoring table: IDENTITY_NUC

Searched: 2002476 seqs, 762712212 residues

Database :   EST:*
                1: em_est1:*
                2: em_est2:*
                3: em_est3:*
                4: em_est4:*
                5: em_est5:*
                6: em_est6:*
                7: em_est7:*
                8: em_est8:*
                9: em_est9:*
               10: gb_est1:*
               11: gb_est2:*
               12: gb_est3:*
               13: gb_est4:*
               14: gb_est5:*
               15: gb_est6:*
               16: gb_est7:*
               17: gb_est8:*
               18: gb_est9:*
               19: gb_est10:*
               20: gb_est11:*
               21: gb_est12:*
               22: gb_est22:*
               23: gb_est13:*
               24: gb_est14:*
               25: gb_est15:*
               26: gb_est16:*
               27: gb_est17:*
               28: gb_est18:*
               29: gb_est19:*
               30: gb_est20:*
               31: gb_est21:*
               32: em_est10:*
               33: em_est11:*
               34: em_est12:*
               35: em_est13:*
               36: em_est14:*
               37: em_est15:*
               38: em_est16:*
               39: em_est17:*
               40: em_est18:*
               41: em_est19:*

No.	c						Accession	Description
3	c	6.8	41.8	419	31	AU033391	AU033391	Dictyoste
4		6.7	41.4	358	27	AA634656	AA634656	ab26b06.r
5	c	6.7	41.4	591	31	AU033325	AU033325	Dictyoste
6		6.6	40.6	554	19	C23753	C23753	Dictyosteli
7		6.4	39.8	574	28	C91220	C91220	Dictyosteli
8		6.4	39.4	312	31	AI163084	AI163084	A03lp65u
9		6.4	39.4	397	31	AU034056	AU034056	Dictyoste
10		6.3	39.2	472	12	N39467	N39467	yw62d10.r1
11		6.3	39.2	552	28	C91403	C91403	Dictyosteli
12		6.3	39.2	480	28	C93684	C93684	Dictyosteli
13	c	6.3	39	402	11	R99317	R99317	yq71h03.s1
14		6.3	39	546	17	C22974	C22974	Dictyoste
15		6.3	39	246	31	AU033454	AU033454	Dictyoste
16		6.3	39	617	31	AU033655	AU033655	Dictyoste
17		6.3	38.8	198	31	AU034673	AU034673	Dictyoste
18		6.2	38.6	466	20	C25660	C25660	Dictyosteli
19		6.2	38.4	408	31	AU033476	AU033476	Dictyoste
20		6.2	38.2	249	28	C94333	C94333	Dictyosteli
21		6.2	38.2	526	31	AU033912	AU033912	Dictyoste
22		6.2	38	627	28	C91439	C91439	Dictyoste
23		6.1	38	294	30	AU034848	AU034848	Dictyoste
24		6.1	37.6	585	20	C24619	C24619	Dictyosteli
25		6.1	37.6	606	25	C84091	C84091	Dictyosteli
26		6.1	37.6	427	28	C91312	C91312	Dictyosteli
27		6.1	37.4	287	28	C94188	C94188	Dictyosteli
28		6.1	37.4	539	29	AI070146	AI070146	UI-R-Y0-1
29		6.1	37.4	414	29	AU033929	AU033929	Dictyoste
30		6.0	37	470	26	C84849	C84849	Dictyosteli
31		6.0	36.8	545	25	C83861	C83861	Dictyosteli
32		5.9	36.6	503	29	AI069527	AI069527	BSBmMFSz1
33	c	5.9	36.6	422	16	AA125493	AA125493	mg79d11.r
34		5.9	36.2	553	22	AA528418	AA528418	ne84a12.s
35		5.9	36.2	458	20	AI291213	AI291213	qm16b05.x
36	c	5.9	36.2	630	26	AA744771	AA744771	ny71d07.s
37	c	5.9	36.2	526	26	AA775720	AA775720	zf31d09.s
38	c	5.8	36	528	28	C92028	C92028	Dictyosteli
39		5.8	36	531	28	C93163	C93163	Dictyosteli
40		5.8	35.8	486	20	AA518458	AA518458	vi06h05.r
41		5.8	35.8	533	29	AA710415	AA710415	vt53f08.r
42		5.8	35.8	278	30	AI153808	AI153808	ud49a02.r
43		5.8	35.6	524	28	C91295	C91295	Dictyosteli
44		5.7	35.4	484	20	C25586	C25586	Dictyosteli
45		5.7	35	456	10	R53852	R53852	yi03d01.r1

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	42.4	6.9	303	31	AU034213	AU034213 Dictyoste
2	42.2	6.8	482	28	C91001	C91001 Dictyosteli

ALIGNMENTS

RESULT 1
AU034213
LOCUS   AU034213   303 bp   mRNA   EST   28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLC264, mRNA sequence.
ACCESSION   AU034213
NID   g3799637
KEYWORDS   EST.
SOURCE   Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM   Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 303)
AUTHORS   Urushihara,H.
TITLE   Developmental cDNA in Dictyostelium discoideum
JOURNAL   Published Only in Database (1998) In press
REFERENCE   2 (bases 1 to 303)
AUTHORS   Urushihara,H.
TITLE   Direct Submission
JOURNAL   Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT   PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES   Location/Qualifiers

```
source          1..303
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="SL"
                /dev_stage="slug"
BASE COUNT        172 a      67 c      12 g      52 t
ORIGIN

Query Match             6.9%;  Score 42.4;  DB 31;  Length 303;
Best Local Similarity  58.9%;  Pred. No. 0.23;
Matches  73;  Conservative  0;  Mismatches  51;  Indels  0;  Gaps  0;

Qy      9 catggatagaaacagttcaacagtagtacttcaaaaacaaaaagcatcatggcaagaca  68
          || |||||||| ||||||| || |||| |||||||||||| ||||||| | |||||||
Db     54 CAACAACAACAACCACTACAACAACCACTACAACAACAACAACAACAACAACAACACAA  113

Qy     69 gatggagtacttgaataacaacaatggcgacaataacaacaacaacaataatgttacaagctc  128
          | | || ||||| |  |||||||| | ||||||||| ||| ||| | ||| |||||||||
Db    114 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATATTATTACCAA  173

Qy    129 atta  132
          | ||
Db    174 ACTA  177


RESULT  2
C91001
LOCUS       C91001       482 bp    mRNA             EST       15-APR-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SSJ551, mRNA sequence.
ACCESSION   C91001
NID         g3060367
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 482)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 482)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES            Location/Qualifiers
source          1..482
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="SS"
                /dev_stage="slug"
BASE COUNT        232 a      98 c      51 g     101 t
ORIGIN

Query Match             6.8%;  Score 42.2;  DB 28;  Length 482;
Best Local Similarity  61.3%;  Pred. No. 0.29;
Matches  68;  Conservative  0;  Mismatches  43;  Indels  0;  Gaps  0:

Qy      9 catggatagaaacagttcaacagtagtacttcaaaaacaaaaagcatcatggcaagaca  68
          || |||||||| ||||||| || |||| |||||||||||| ||||||| | |||||||
Db    368 CAGCAACAGCAACAATAACACACACAACAACAACAACAACAACAACAACAACAACAACAA  427

Qy     69 gatggagtacttgaataacaacaatggcgacaataacaacaacaacaataatgt  119
          | | || ||||| |  |||||||| | ||||||||| ||| ||| | |||
Db    428 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATAT  478


RESULT  3
AU033391
LOCUS       AU033391     419 bp    mRNA             EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA713, mRNA sequence.
ACCESSION   AU033391
NID         g3798815
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 419)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 419)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES            Location/Qualifiers
source          1..419
                /organism="Dictyostelium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="SL"
                /dev_stage="slug"
BASE COUNT        248 a      91 c      24 g      56 t
ORIGIN

Query Match             6.8%;  Score 41.8;  DB 31;  Length 419;
Best Local Similarity  61.5%;  Pred. No. 0.35;
Matches  67;  Conservative  0;  Mismatches  42;  Indels  0;  Gaps  0:

Qy      9 catggatagaaacagttcaacagtagtacttcaaaaacaaaaagcatcatggcaagaca  68
          || |||||||| ||||||| || |||| |||||||||||| ||||||| | |||||||
Db    191 CAGCAACAAGCACAATAACACACACAACAACAACAACAACAACAACAACAACAACAACAA  250

Qy     69 gatggagtacttgaataacaacaatggcgacaataacaacaacaacaataat  117
          | | || ||||| |  |||||||| | ||||||||| ||| ||| | |||
Db    251 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAT  299


RESULT  4
AA634656/c
LOCUS       AA634656     358 bp    mRNA             EST       06-MAR-1998
DEFINITION  ab26b06.r1 Stratagene lung (#937210) Homo sapiens cDNA clone 841907
            5' similar to TR:G510703 G510703 LAMININ S B3 CHAIN PRECURSOR. [1]
            ;, mRNA sequence.
ACCESSION   AA634656
NID         g2557870
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 358)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
```

```
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               This clone is available royalty-free through LLNL ; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Insert Length: 759    Std Error: 0.00
               Seq primer: -28m13 revl ET from Amersham
               High quality sequence stop: 348.
FEATURES       Location/Qualifiers
     source    1..358
               /organism="Homo sapiens"
               /note="Organ: lung; Vector: pBluescript SK-; Site_1:
               EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
               Oligo dT normal lung.  Average insert size: 1.0 kb;
               Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
               3', -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
               /db_xref="taxon:9606"
               /clone="841907"
               /clone_lib="Stratagene lung (#937210)"
               /sex="male"
               /dev_stage="72 years"
               /lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT     74 a     76 c    106 g    102 t
ORIGIN

Query Match              6.7%;   Score 41.4;   DB 27;   Length 358;
Best Local Similarity   56.1%;   Pred. No. 0.43;
Matches    78;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     13 gatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagacagatg 72
          ||| ||||||| |||| ||||||| ||||| ||||||||||||| ||||||| |||
Db    317 GATGGAAAGAATTCAAACCCAATCTACCCCCAACACACAAAGCGGCAGATGAGTG 258

Qy     73 gagtacttgaataaccaacaatggcgacaataacaacaacaataatgttacaagctcatta 132
          |||| || ||||||| ||| |||||||||||| || ||||| |||||| | ||||
Db    257 GGGCAACGGGATGAGAAAGCTGTGAGCAACACTTACAGGTAGCATAGTAGAGCACCGCGCACAT 198

Qy    133 cgagataatatggaaatg 151
          || ||||| ||||| ||
Db    197 TGAATATGATCACGGATCTG 179

RESULT 5
AU033325
LOCUS       AU033325      591 bp    mRNA             EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA616, mRNA sequence.
ACCESSION   AU033325
NID         g3798749
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 591)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 591)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES    Location/Qualifiers
     source  1..591
            /organism="Dictyostelium discoideum"
            /strain="AX4"
            /db_xref="taxon:44689"
            /clone_lib="SL"
            /dev_stage="slug"
BASE COUNT    234 a    115 c     77 g    162 t     3 others
ORIGIN

Query Match              6.7%;   Score 41.4;   DB 31;   Length 591;
Best Local Similarity   55.1%;   Pred. No. 0.48;
Matches    81;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      9 catggtagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagaca 68
          |||||||||| ||||||| |||| ||||||| ||||| |||||| ||| || ||
Db    375 CAACAACAACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 434

Qy     69 gatgggtacttgaataacaacaatggcgacaataacaacaacaacaacaataatgttacaagctc 128
          ||||| | || ||||||| ||| |||||||||||| ||||||||||||||| || ||||
Db    435 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAAAACTACAACC 494

Qy    129 attacgagataattatggaaatgaaga 155
          | ||| |||||| || ||||| ||||
Db    495 ACAACCTCAACCTCATGTACATGTAGA 521

RESULT 6
C23753/c
LOCUS       C23753        554 bp    mRNA             EST       22-APR-1997
DEFINITION  Dictyostelium discoideum gamete cDNA, clone FC-AF03, mRNA sequence.
ACCESSION   C23753
NID         g2073265
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:KAX3) Sexually competent Gamete
            cDNA to mRNA, clone_lib:FC.
ORGANISM    Dictyostelium discoideum
            Eukaryota; mitochondrial eukaryotes; Dictyostelida;
            Dictyostelium.
REFERENCE   1  (bases 1 to 554)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (04-APR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1 Tennohdai, Tsukuba, Ibaraki 305, Japan
            (E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
            Fax:0298-53-6614)
REFERENCE   2  (sites)
AUTHORS     Kobayashi,A., Shimizu,H. and Urushihara,H.
TITLE       Sexual cDNA in CSM
JOURNAL     Unpublished (1997)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES    Location/Qualifiers
     source  1..554
            /organism="Dictyostelium discoideum"
            /strain="KAX3"
            /db_xref="taxon:44689"
            /cell_type="Gamete"
            /clone_lib="FC"
            /dev_stage="Sexually competent"
BASE COUNT    117 a     36 c    123 g    278 t
ORIGIN

Query Match              6.6%;   Score 40.6;   DB 19;   Length 554;
Best Local Similarity   62.1%;   Pred. No. 0.77;
Matches    64;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     14 atagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaagacagatgg 73
          | | ||||||| |||||||| ||||||| || | ||||| ||||||||||| | ||
Db    172 ACAACAACAGTATCACCAACAACAAGACAGCAACAACAATATTATCAACAACAACAACAACA 113

Qy     74 agtacttgaataacaacaatggcgacaataacaacaacaacaacaacaacaata 116
          || ||| ||||||| ||| ||||||| || ||||||||||||||| |||||
Db    112 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 70
```

```
RESULT   7
         C91220
LOCUS      C91220      574 bp     mRNA             EST       15-APR-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SSJ834, mRNA sequence.
ACCESSION  C91220
NID        g3060586
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE     Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
           clone_lib:SS.
           Dictyostelium discoideum
ORGANISM   Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE  1 (bases 1 to 574)
AUTHORS    Urushihara,H.
TITLE      Developmental cDNA in Dictyostelium discoideum
JOURNAL    Published Only in DataBase (1998) In press
REFERENCE  2 (bases 1 to 574)
AUTHORS    Urushihara,H.
TITLE      Direct Submission
JOURNAL    Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
           Urushihara, University of Tsukuba, Institute of Biological
           Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
           (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
           Fax:+81-0298-53-6614)
COMMENT    PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES   Location/Qualifiers
  source     1..574
             /organism="Dictyostelium discoideum"
             /strain="AX4"
             /db_xref="taxon:44689"
             /clone_lib="SS"
             /dev_stage="slug"
BASE COUNT     258 a     63 g     176 t     1 others
ORIGIN

Query Match             6.4%;  Score 39.8;  DB 28;  Length 574;
Best Local Similarity  60.7%;  Pred. No. 1.2;
Matches   65;  Conservative   0;  Mismatches   42;  Indels   0;  Gaps   0;

Qy   38 acttcaaaaacaaaagcatcatgcgcaagacagatggagtacttgaataacaacaatggcg  97
        || ||||||||| || |||| ||| || ||||| ||||| || || |||||||| ||||||
Db  386 AATTGAAAAACAATATAATAATAGAAAGCAACAACAACAACAACAACAACAACAACAACACA 445

Qy   98 acaataacaacaacaataatgttacaagctcattacgaggataattat 144
        ||| ||||||||||| ||        | | | ||||||| |||||||
Db  446 ACACAACAACAACAACAACAACAACAACAACAACAACAACAAAATAATTAT 492

RESULT   8
         AI163084
LOCUS      AI163084      312 bp     mRNA             EST       03-DEC-1998
DEFINITION A031p65u Hybrid aspen plasmid library Populus tremula x Populus
           tremuloides cDNA 5', mRNA sequence.
ACCESSION  AI163084
NID        g3854369
KEYWORDS   EST.
SOURCE     Populus tremula x Populus tremuloides.
ORGANISM   Populus tremula x Populus tremuloides
           Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
           Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta;
           Magnoliophyta; eudicotyledons; Rosidae; Violales; Salicaceae;
           Populus.
REFERENCE  1 (bases 1 to 312)
AUTHORS    Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A.,
           Holmberg,A., Amini,B., Bhalerao,R., Larsson,M., Villarroel,R., Van
           Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W.,
           Gustafsson,P., Uhlen,M., Sundberg,B. and Lundeberg,J.
TITLE      Gene discovery in the wood-forming tissues of poplar: Analysis of
           5,692 expressed sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE    99007314
COMMENT    Contact: Sterky F
```

```
           Department of Biotechnology
           Royal Institute of Technology (KTH)
           Teknikringen 34, S-100 44 STOCKHOLM, Sweden
           Tel: +46 8 790 8287
           Fax: +46 8 24 54 52
           Email: fredrik@biochem.kth.se
           PCR PRimers
           FORWARD: AAAGGGGATGTGTCGCAAGGCG
           BACKWARD: GCTTCCGGCTCGTATGTTGTGTG
           Seq primer: CGTTGTAAAACGACGGCCAG
           High quality sequence stop: 312.
FEATURES   Location/Qualifiers
  source     1..312
             /organism="Populus tremula x Populus tremuloides"
             /note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
             Cambial region tissues, including developing xylem, the
             meristematic cambial zone and the developing phloem,
             was harvested from 1.5 m actively growing trees.
             cDNA was prepared and cloned into lambda g22a. DNA was
             isolated and subcloned into pBluescript SK using SalI and
             NotI restriction enzymes."
             /db_xref="taxon:47564"
             /clone_lib="Hybrid aspen plasmid library"
             /tissue_type="Cambial region"
             /dev_stage="1.5 m actively growing tree"
             /lab_host="E.coli"
BASE COUNT     113 a     66 c     71 g     62 t
ORIGIN

Query Match             6.4%;  Score 39.4;  DB 23;  Length 312;
Best Local Similarity  65.2%;  Pred. No. 1.4;
Matches   58;  Conservative   0;  Mismatches   31;  Indels   0;  Gaps   0;

Qy  187 ccaagatcttatacatgtagctttgtaaaagggaattttagatcgtgctcaagctccttggt 246
        | ||| ||  |||||| |||| ||| |||| |      |   ||||| ||||||||
Db  200 CAAAGGGTTTCTCATGCAACTATTGCCAAAGAAAATTTTATAGCTCACAAGCGCTTGGA 259

Qy  247 ggacacatgaatgttcatagaagagatag 275
        |||||| | || ||| || ||| ||||| |
Db  260 GGACACCAAAATGCCATAAAAGAGAAAG 288

RESULT   9
         AU034056
LOCUS      AU034056      397 bp     mRNA             EST       28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLB863, mRNA sequence.
ACCESSION  AU034056
NID        g3799480
KEYWORDS   EST.
SOURCE     Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
           clone_lib:SL.
           Dictyostelium discoideum
ORGANISM   Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE  1 (bases 1 to 397)
AUTHORS    Urushihara,H.
TITLE      Developmental cDNA in Dictyostelium discoideum
JOURNAL    Published Only in DataBase (1998) In press
REFERENCE  2 (bases 1 to 397)
AUTHORS    Urushihara,H.
TITLE      Direct Submission
JOURNAL    Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
           Urushihara, University of Tsukuba, Institute of Biological
           Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
           (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
           Fax:+81-0298-53-6614)
COMMENT    PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES   Location/Qualifiers
  source     1..397
             /organism="Dictyostelium discoideum"
             /strain="AX4"
             /db_xref="taxon:44689"
             /clone_lib="SL"
```

Qy 265 agagagatagagccatttgagacaatcaccacctgagatattaataggtattcctt 324
         ||-|||-||| ||| |||| || ||||||| |||| |||||| |||||| |||||||
Db 270 AGTGGGATGGATTCATGTCGATACANTCCCCATGCCCTTGACCTTCTCTGGCATTCTCCT 329

Qy 325 ctaaacctaatcttgaaccaaacctaacttttaccctagcataaccctagtttttca 384
         || ||| |||||||||||| ||||||||||| |||||||| |||||| ||||| ||
Db 330 GTGCTCTGACAAACTGAGCCAGCCTTTTAGATCTACATGAATAACAAACTATTTTACCA 389

Qy 385 agaaaattcccaccttttgaaatgaggaaattaggaa 421
         |||| ||| || |||||||| |||||||||| |||| ||
Db 390 AGAAAAATCTCAGCTTGCTTACTGNCTTACTGNCTTAATTAAAAA 426

RESULT 11
C91403
LOCUS   C91403   552 bp   mRNA   EST   15-APR-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SSK172, mRNA sequence.
ACCESSION   C91403
NID   g3060769
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE   Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 552)
  AUTHORS   Urushihara,H.
  TITLE   Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Published Only in DataBase (1998) In press
REFERENCE   2 (bases 1 to 552)
  AUTHORS   Urushihara,H.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-MAR-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT   PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES   Location/Qualifiers
     source   1..552
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="slug"
BASE COUNT   241 a   127 c   42 g   139 t   3 others
ORIGIN

Query Match   6.3%; Score 39.2; DB 28; Length 552;
Best Local Similarity   60.2%; Pred. No.1.8;
Matches   65; Conservative   0; Mismatches   43; Indels   0; Gaps   0;

Qy 9 catggtagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagaca 68
       || ||||| |||| |||| ||| | |||||||| |||||||||| || || | || ||
Db 353 CAACAACAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATAA 412

Qy 69 gatggagtacttgaataacaacaatggcgacaataacaacaacaataa 116
       ||| || | |||| ||||| || |||| ||| || |||| || ||
Db 413 CAATAACAACAACAATAGCAACAATATCAACAATGATAACAACAACAA 460

RESULT 12
C93684
LOCUS   C93684   480 bp   mRNA   EST   06-JUN-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SSL603, mRNA sequence.
ACCESSION   C93684
NID   g3192076
KEYWORDS   EST; EST(expressed sequence tag).
SOURCE   Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SS.
  ORGANISM Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1 (bases 1 to 480)

BASE COUNT   207 a   94 c   26 g   69 t   1 others
ORIGIN

Query Match   6.4%; Score 39.4; DB 31; Length 397;
Best Local Similarity   62.9%; Pred. No.1.5;
Matches   61; Conservative   0; Mismatches   36; Indels   0; Gaps   0;

Qy 16 agaaacagtttcaacagtagtacttcaaaaacaaaagcatcatggcaagacagatggag 75
        | ||| |||||||||| | || |||| ||||||||| |||| || || ||
Db 123 ACAATCAACTTCAACGACAACATTTCAACAACAACTTCAACAACAACAACCATTTTC 182

Qy 76 tacttgaataacaacaatggcgacaataacaacaaca 112
        || ||||||||||||
Db 183 AACAACAACAACAACAACAACAACAGCAACAACA 219

RESULT 10
N39467
LOCUS   N39467   472 bp   mRNA   EST   19-JAN-1996
DEFINITION yw62d10.r1 Homo sapiens cDNA clone 256819 5'.
ACCESSION   N39467
NID   g1162674
KEYWORDS   EST.
SOURCE   human clone=256819 primer=T7 library=Soares placenta 8to9weeks
            2NbHP8to9W vector=pT7T3D (Pharmacia) with a modified polylinker
            host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI two
            placentae: one from 8 weeks and another from 9 weeks post
            conception. 1st strand cDNA was primed with a Not I - oligo(dT)
            primer [5'-TGTTACCAATCTGAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI adapters
            (Pharmacia), digested with Not I and cloned into the Not I and Eco
            RI sites of a modified pT7T3 vector (Pharmacia). Library
            constructed by Bento Soares and M. Fatima Bonaldo.
  ORGANISM Homo sapiens
            Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
            Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
            Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
            Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 472)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE   The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 320
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES   Location/Qualifiers
     source   1..472
                     /organism="Homo sapiens"
                     /clone="256819"
     mRNA   <1..>472
                     /clone="256819"
BASE COUNT   126 a   94 c   73 g   177 t   2 others
ORIGIN

Query Match   6.3%; Score 39.2; DB 12; Length 472;
Best Local Similarity   52.9%; Pred. No.1.7;
Matches   83; Conservative   0; Mismatches   74; Indels   0; Gaps   0;

```
AUTHORS      Yoshida,M.
TITLE        Developmental cDNA in Dictyostelium discoideum
JOURNAL      Published Only in DataBase (1998) In press
REFERENCE    2  (bases 1 to 480)
AUTHORS      Yoshida,M.
TITLE        Direct Submission
JOURNAL      Submitted (11-MAY-1998) to the DDBJ/EMBL/GenBank databases.
             Motonobu Yoshida, Research Institute of Food Science, Kinki
             University: Nakamachi 3327, Nara 631-0052, Japan
             (E-mail:yoshida@nara.kindai.ac.jp, Tel:81-742-43-1511)
COMMENT      PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..480
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SS"
                     /dev_stage="slug"
BASE COUNT        187 a       119 c       34 g       140 t
ORIGIN

Query Match                  6.3%;   Score 39.2;   DB 28;   Length 480;
Best Local Similarity   60.2%;   Pred. No. 1.7;
Matches   65;   Conservative    0;   Mismatches   43;   Indels    0;   Gaps    0;

Qy      9 catggtagaaacagttcaacagtagtacttcaaaaacaaagcatcatggcaagaca  68
          |||||| | |||||| |||| ||||| |||||||||||||||||||| || | |||
Db    249 CATCACAACAACTACAACAACAACAACTACTACAACAACAACAACAACAACAACAA 308

Qy     69 gatggagtacttgaataacaacaatggcgcacaataacaacaacaataa 116
          | || |||||||||| ||||||||| |||||||||||||||||||||||||
Db    309 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA 356

RESULT 13
R99317/c
LOCUS       R99317        402 bp   mRNA            EST          14-SEP-1995
DEFINITION  yg71h03.s1 Homo sapiens cDNA clone 201269 3'.
ACCESSION   R99317
NID         g985918
KEYWORDS    EST.
SOURCE      human clone=201269 primer=Promega -21m13 library=Soares fetal liver
             spleen 1NFLS vector=pT7T3D (Pharmacia) with a modified polylinker
             host=DH10B (ampicillin resistant) Rsite1=Pac I Rsite2=Eco RI Liver
             and spleen from a 20 week-post conception male fetus. 1st strand
             cDNA was primed with a Pac I - oligo(dT) primer [5'-
             AACTGGAAGAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded
             cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
             I and cloned into the Pac I and Eco RI sites of the modified pT7T3
             vector. Library went through one round of normalization. Library
             constructed by Bento Soares and M.Fatima Bonaldo.
ORGANISM    Homo sapiens
             Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
             Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
             Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
             Eutheria; Archonta; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 402)
AUTHORS     Hiller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
             Holman,M., Hultman,M., Hultman,M., Le.M., Lennon,G., Marra,M.,
             Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
             Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
             Wilson,R.
TITLE       The WashU-Merck EST Project
JOURNAL     Unpublished (1995)
COMMENT     Contact: Wilson RK
             WashU-Merck EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
```

```
             High quality sequence stops: 337
             Source: IMAGE Consortium, LLNL
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES             Location/Qualifiers
     source          1..402
                     /organism="Homo sapiens"
                     /clone="201269"
     mRNA            <1..>402
BASE COUNT        126 a        70 c        75 g       131 t
ORIGIN

Query Match                  6.3%;   Score 39;   DB 11;   Length 402;
Best Local Similarity   52.8%;   Pred. No. 1.8;
Matches   84;   Conservative    0;   Mismatches   75;   Indels    0;   Gaps    0;

Qy    265 agaggagatagagccatttgagacaatcaccacctagagatattaataggtattcctctt 324
          |||| ||| |||||  ||| |||| | |  | ||| || || ||| || || | ||||
Db    325 AGAAAGAAAGTACCCTGTTCCAAAAGCACTGCCTCACATTGTCAGTAAATTCTTTCTG 266

Qy    325 ctaaaccttaatcttgaaccaaaccctaacttaccctagtcataacccctagttttca 384
          ||| | ||| |||| |||||||  | ||| || |||||| |||| ||||| || |||
Db    265 TTTAGCTGAAAAAGAAAAAACCATTCTGTTAATTTTCCATTGAAATAGACTCATATATGA 206

Qy    385 agaaattcccacctttgaaatgaggaaattaggaaaa 423
          ||||  |||| | |  |  ||||||  || ||||||
Db    205 AGACAATACTTAAACTTGTAATCATGCTGGTTGAACA 167

RESULT 14
C22974
LOCUS       C22974        546 bp   mRNA
DEFINITION  Dictyostelium discoideum gamete cDNA, clone FC-AL20, mRNA sequence.
ACCESSION   C22974                                              EST          27-MAR-1997
NID         g1922123
KEYWORDS    EST; EST(expressed sequence tag).
SOURCE      Dictyostelium discoideum (strain:KAX3) Sexually mature Gamete cDNA
             to mRNA, clone_lib:FC.
             Dictyostelium discoideum
ORGANISM    Eukaryotae; mitochondrial eukaryotes; Dictyostellida;
             Dictyostelium.
REFERENCE   1  (bases 1 to 546)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAR-1997) to the DDBJ/EMBL/GenBank databases. Hideko
             Urushihara, University of Tsukuba, Institute of Biological
             Sciences: Ten-noudai 1-1, Tsukuba-shi, Ibaraki-ken 305, Japan
             (E-mail:d402hu@sakura.cc.tsukuba.ac.jp, Tel:0298-53-4664,
             Fax:0298-53-6614)
REFERENCE   2  (sites)
AUTHORS     Shimizu,H. and Urushihara,H.
TITLE       Sexual-cDNA in CSM
JOURNAL     Unpublished (1997)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..546
                     /organism="Dictyostelium discoideum"
                     /strain="KAX3"
                     /db_xref="taxon:44689"
                     /cell_type="Gamete"
                     /clone_lib="FC"
                     /dev_stage="Sexually mature"
BASE COUNT        317 a       110 c        33 g        85 t        1 others
ORIGIN

Query Match                  6.3%;   Score 39;   DB 17;   Length 546;
Best Local Similarity   59.5%;   Pred. No. 2;
Matches   66;   Conservative    0;   Mismatches   45;   Indels    0;   Gaps    0;

Qy      6 aggcatggatagaaacagttcaacagtagtacttcaaaaacaaaagcatcatggcaag 65
          || || ||||| ||||||||||||||||||||||| ||||||| || ||||||||||||
```

```
Db      146 AGCAATTATTATCAAAAAATTGAACAACAACAACAACAACAACAACAACAA 205

Qy       66 acagatggagtacttgaataacaatggcgacaataacaacaacaataa 116
             ||-| | | || ||||||| | |||| |||||||||| ||
Db      206 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 256


RESULT 15
LOCUS       AU033454        246 bp    mRNA         EST       28-OCT-1998
DEFINITION  Dictyostelium discoideum slug cDNA, clone SLA845, mRNA sequence.
ACCESSION   AU033454
NID         g3798878
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
            clone_lib:SL.
ORGANISM    Dictyostelium discoideum
            Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE   1  (bases 1 to 246)
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2  (bases 1 to 246)
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664,
            Fax:+81-0298-53-6614)
COMMENT     PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES             Location/Qualifiers
     source          1..246
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="SL"
                     /dev_stage="slug"
BASE COUNT      142 a      67 c      9 g       28 t
ORIGIN

Query Match         6.3%;   Score 39;  DB 31;  Length 246;
Best Local Similarity  59.5%;   Pred. No. 1.6;
Matches  66;   Conservative   0;  Mismatches  45;  Indels   0;  Gaps   0;

Qy        9 catggatagaaacagttcaacagtaagtacttcaaaaacaaaagcatcatggcaagaca 68
            || | | |||| |||| | || ||||  || | |  ||| ||||| | || || | |  |
Db       22 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 81

Qy       69 gatggagtacttgaataacaacaatggcgacaataacaacaacaacaataatgt 119
            || -| | || ||||||| | |||| |||||||||| || ||
Db       82 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACGT 132


Search completed: July 15, 1999, 12:07:22
Job time: 3284 sec
```

```
*****************************************************

   |\/| |  |  |‾‾| |‾‾| |‾‾| |  |
   |  | |‾‾|  |‾‾  |‾‾| |    |‾‾|
   |  | |  |  |    |  | |  | |  |  (TM)

*****************************************************

       Release 3.1A John F. Collins, Biocomputing Research Unit.
       Copyright (c) 1993-1998 University of Edinburgh, U.K.
                 Distribution rights by Oxford Molecular Ltd
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jul 14 17:26:36 1999; MasPar time 15.20 Seconds
                                          288.149 Million cell updates/sec

Tabular output not generated.

Title: >US-09-156-580-2
Description: (1-206) from US09156580A.pep
Perfect Score: 1495
Sequence: 1 MAGMDRNSFNSKYFKNKSIM.........IGLISESKEDDLELRLGST 206

Scoring table: PAM 150
                 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database: a-geneseq35
                 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
                 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
                 14:part14 15:part15 16:part16 17:part17 18:part18
                 19:part19 20:part20 21:part21 22:part22 23:part23
                 24:part24 25:part25 26:part26 27:part27 28:part28
                 29:part29 30:part30 31:part31 32:part32 33:part33
                 34:part34 35:part35 36:part36 37:part37 38:part38
                 39:part39

Statistics: Mean 31.942; Variance 133.844; scale 0.239

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES

Result   %   Query
No.   Score Match Length DB ID   Description   Pred. No.
---- ----- ----- ------ -- -------- -----------   ---------
  1   119   8.0   227 20 W01604   Arabidopsis STZ polyp   1.81e-01
  2   115   7.7   2391 10 R55694   Carbamoyl-phosphate-s   3.78e-01
  3   114   7.6   570 26 W36002   Human Fchd531 gene pr   4.54e-01
  4   111   7.4   31 37 W75038   Fragment of human sec   7.85e-01
  5   111   7.4   183 37 W75040   Fragment of human sec   7.85e-01
  6   111   7.4   203 37 W74793   Human secreted protei   7.85e-01
  7   101   6.8   364 20 W01457   E. coli O157:H7 F12 r   4.70e+00
  8   97   6.5   88 38 W87557   A nucleic acid bindin   9.48e+00
  9   97   6.5   88 39 W84299   An anti-HIV zinc fing   9.48e+00
 10   94   6.3   1060 30 W23975   Homo sapiens 20q13 am   1.59e+01
 11   92   6.2   423 17 R91305   Transcription factor-   2.24e+01
 12   92   6.2   706 12 R68743   BCL-6 zinc finger pro   2.24e+01
 13   93   6.2   803 37 W81756   Myc-binding zinc-fing   1.89e+01
 14   91   6.1   547 17 R95242   HIC-1 polypeptide.   2.66e+01
 15   91   6.1   957 22 W20841   H. pylori transmembra   2.66e+01
 16   91   6.1   2496 26 W36093   Mutant YLR087c protei   2.66e+01
 17   91   2942 25 W22050   Saccharomyces cerevis   2.66e+01
 18   89   6.0   92 18 R92699   zinc finger protein z   3.73e+01
 19   88   6.0   1719 16 R92100   Human RIZ allele D283   3.73e+01
 20   88   5.9   343 16 R77656   Human zinc finger pro   4.42e+01
 21   88   5.9   566 37 W82559   A. thaliana homologou   4.42e+01
 22   88   5.9   675 17 R94903   Human lastin.   4.42e+01
 23   88   5.9   1706 16 R92103   Rat RIZ.   4.42e+01
 24   87   5.8   59.12   R63135 Egr-1 nuclear localiz   5.22e+01
 25   87   5.8   89 11 R63136   Egr-1 polynucleotide   5.22e+01
 26   87   5.8   91 39 W84456   zif268 three finger p   5.22e+01
 27   87   5.8   91 15 R83492   zif268 zinc finger do   5.22e+01
 28   87   5.8   153 39 W84391   zif268-Jun construct.   5.22e+01
 29   87   5.8   153 39 W84392   zif268-Fos construct.   5.22e+01
 30   87   5.8   153 15 R83488   zif268-Fos zinc finge   5.22e+01
 31   87   5.8   153 15 R83487   zif268-Jun zinc finge   5.22e+01
 32   87   5.8   181 39 W84394   The zif268-zif268 con   5.22e+01
 33   87   5.8   181 15 R83490   zif268-zif268 zinc fi   5.22e+01
 34   86   5.8   387 37 W76987   Human EGR-3 protein.   6.17e+01
 35   87   5.8   455 37 W76986   Human EGR-2 protein.   5.22e+01
 36   87   5.8   456   1 P93114   EGR2.   5.22e+01
 37   87   5.8   456 41 R63130   Human Egr-1 protein.   5.22e+01
 38   87   5.8   496 37 W76984   Mouse Egr-1 protein.   5.22e+01
 39   87   5.8   500 32 W37948   Amino acid sequence o   5.22e+01
 40   86   5.8   507 22 W15574   Mouse Aiolos polypept   6.17e+01
 41   87   5.8   533 11 R63129   Mouse Egr-1 clone OC3   5.22e+01
 42   87   5.8   533   1 P93113   Egr-1.   5.22e+01
 43   87   5.8   543 37 W76985   Human EGR-1 protein.   5.22e+01
 44   87   5.8   543   2 R24022   Human promyelo-leukae   5.22e+01
 45   86   5.8   791 17 R95566   N. meningitidis serot   6.17e+01
```

                              ALIGNMENTS

```
RESULT  1
ID   W01604 standard; Protein; 227 AA.
AC   W01604;
DT   16-APR-1997  (first entry)
DE   Arabidopsis STZ polypeptide.
KW   STZ; STO: salt tolerance; transgenic plant; yeast.
OS   Arabidopsis thaliana.
PN   WO9639020-A1.
PD   12-DEC-1996.
PF   31-MAY-1996;  U08095.
PR   06-JUN-1995; US-471717.
PA   (REGC ) UNIV CALIFORNIA.
PI   Gasser CS,  Lippuner V;
DR   WPI; 97-042725/04.
DR   N-PSDB; T58281.
PT   Nucleic acids encoding STZ and STO polypeptide(s) which confer salt
PT   tolerance - useful for producing transgenic plants or other
PT   organisms, e.g. yeast, able to grow under saline conditions
PS   Claim 2; Page 22; 31pp; English.
CC   The STZ (W01604) and STO (W01605) polypeptides of Arabidopsis
CC   thaliana have the ability to confer tolerance to saline conditions
CC   on plants, yeast, fungi and other organisms. They can be obtd. by
CC   expression of STZ (T58281) and STO (T58282) cDNA clones that were
CC   isolated from an A. thaliana library by complementation of a salt
CC   tolerance phenotype in yeast.  Recombinant expression cassettes
CC   can be used to express the STZ and STO polypeptides in transgenic
CC   plants, allowing the plants to grow under saline conditions.
SQ   Sequence   227 AA;

Query Match             8.0%;   Score 119; DB 20; Length 227;
Best Local Similarity  55.6%;   Pred. No. 1.81e-01;
  Matches   15;  Conservative   4;  Mismatches   8;  Indels   0;  Gaps   0;

Db     79 sykcsvcdktfssyqalgghkashrkn 105
          || || | |||| || | | ||::: || |:::
Qy     65 SYTCSFCKREFRSAQALGGHMNVHRRD 91

RESULT  2
ID   R55694 standard; Protein; 2391 AA.
```

```
AC   R55694;
DT   06-DEC-1994 (first entry)
DE   Carbamoyl-phosphate-synthetase II.
KW   Carbamoyl-phosphate-synthetase II; CPSII; psCPSII gene;
KW   malaria.
OS   Plasmodium falciparum.
FH   Key                 Location/Qualifiers
FT   domain              1..690
FT                       /note= "glutamine-amidotransferase domain"
FT   domain              1..270
FT                       /note= "structural subdomain"
FT   peptide             271..482
FT                       /note= "insert sequence"
FT   domain              483..690
FT                       /note= "glutaminase subdomain"
FT   domain              691..2391
FT                       /note= "carbamoyl-phosphate-synthase domain"
FT   domain              691..1254
FT                       /note= "ATP binding subdomain CPSa"
FT   peptide             1255..1857
FT                       /note= "insert sequence"
FT   domain              1858..2391
FT                       /note= "ATP binding subdomain CPSb"
PN   WO9412643-A.
PD   09-JUN-1994.
PF   02-DEC-1993; AU0617.
PR   03-DEC-1992; AU-006206.
PR   16-DEC-1992; AU-006380.
PA   (UNIX ) UNISEARCH LTD.
PI   Flores MV, Osullivan WJ, Stewart TS;
DR   WPI: 94-200271/24.
DR   N-PSDB: Q62924.
PT   Nucleic acid encoding carbamoyl phosphate synthetase II -
PT   isolated from Plasmodium falciparum, used to develop prods. for
PT   the treatment of malaria.
PS   Disclosure: Page 6-16; 31pp; English.
CC   The cDNA sequence encoding the carbamoyl-phosphate-transferase II
CC   (CPSII) of Plasmodium falciparum was determined.  The cDNA encodes
CC   a protein that includes 2 insert sequences not found in other CPSII
CC   proteins.  The first separates the putative structural subdomain and
CC   the glutaminase subdomain of the glutamine-amidotransferase subunit
CC   of CPSII, while the second separates 2 ATP binding subdomains of the
CC   CPSII subunit, CPSa and CPSb.
SQ   Sequence 2391 AA;

Query Match             7.7%;  Score 115;  DB 10;  Length 2391;
Best Local Similarity   47.6%;  Pred. No. 3.78e-01;
Matches   20;  Conservative   10;  Mismatches   10;  Indels    2;  Gaps    2;

Db      1159 leqlsfndlkyfkkhgfsdkqiahylsfntsdnnnnnniss 1200
              ::: ||| |||||:: :|: :|: ||: | :|||||||| ||
Qy         4 MDRNSFNS-KYFKNKSIMARQME-YLNNNNGDNNNNNNVTSS 43

RESULT 3
ID   W36002 standard: Protein: 570 AA.
AC   W36002;
DT   03-MAR-1998 (first entry)
DE   Human Fchd531 gene product.
KW   Fchd531 gene; differential expression: endothelial cell; human;
KW   shear stress; cardiovascular disease; atherosclerosis; ischaemia;
KW   reperfusion; hypertension; restenosis; arterial inflammation;
KW   therapy; diagnosis: drug screening; marker.
OS   Homo sapiens.
PN   WO9730065-A1.
PD   21-AUG-1997.
PF   14-FEB-1997; U02291.
PR   13-FEB-1997; US-799910.
PR   16-FEB-1996; US-011787.
PA   (MILL-) MILLENNIUM PHARM INC.
PI   Falb DA;
DR   WPI: 97-424966/39.
DR   N-PSDB: T94467.

PT   New genes differentially expressed in cardiovascular disease - used
PT   for diagnosis, drug screening and treatment of cardiovascular
PT   disease, e.g. atherosclerosis, restenosis, hypertension, etc
PS   Example 7; Fig 1A-1B; 163pp; English.
CC   This protein is encoded by the novel human fchd531 gene (see
CC   T94467) that is down-regulated in endothelial cells subjected to
CC   turbulent and laminar shear stress.  Shear stress is thought to be
CC   responsible for the prevelence of atherosclerotic lesions in areas
CC   of unusual circulatory flow.  The fchd531 gene product has 94%
CC   similarity to the mouse penta zinc finger gene (pzf).  Gene
CC   products were also identified for novel genes fchd540. (see
CC   W36003) and fchd545 (see W36004), which are respectively up- and
CC   down-regulated in endothelial cells subjected to shear stress.
CC   Novel fchd531, fchd540, fchd545, fchd602 and fchd605 genes (see
CC   T94467-71) provide a fingerprint for the study of cardiovascular
CC   diseases, including atherosclerosis, ischaemia/reperfusion,
CC   hypertension, restenosis and arterial inflammation.  Methods are
CC   provided for the diagnosis, monitoring in clinical trials,
CC   screening for therapeutically effective compounds, and treatment of
CC   cardiovascular diseases based on discoveries regarding the
CC   expression patterns of these novel genes.
SQ   Sequence 570 AA;

Query Match             7.6%;                 DB 26;  Length 570;
Best Local Similarity   32.3%;  Pred. No. 4.54e-01;
Matches   10;  Conservative   10;  Mismatches   11;  Indels    0;  Gaps    0;

Db       370 rdyiceycarafksshnlavhrmihtgekpl 400
              |||:|| |||| : :|::|: | :|  :::|
Qy        64 RSYTCSFCKREFRSAQLGGHMNVHRRDRAI 94

RESULT 4
ID   W75038 standard: Protein; 31 AA.
AC   W75038;
DT   25-JAN-1999 (first entry)
DE   Fragment of human secreted protein encoded by gene 64.
KW   Human; secreted protein; testis; tumour; foetal brain tissue;
KW   fusion protein; cancer: central nervous system; seizure;
KW   diagnosis; neurodegenerative disease.
OS   Homo sapiens.
PN   WO9839448-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; U04493.
PR   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   23-MAY-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047581.
```

PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   13-JUN-1997; US-049610.
PR   08-JUL-1997; US-051926.
PR   16-JUL-1997; US-052874.
PR   18-AUG-1997; US-055724.
PR   22-AUG-1997; US-056630.
PR   22-AUG-1997; US-056631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056662.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-056862.
PR   22-AUG-1997; US-056864.
PR   22-AUG-1997; US-056872.
PR   22-AUG-1997; US-056874.
PR   22-AUG-1997; US-056875.
PR   22-AUG-1997; US-056876.
PR   22-AUG-1997; US-056877.
PR   22-AUG-1997; US-056878.
PR   22-AUG-1997; US-056879.
PR   22-AUG-1997; US-056880.
PR   22-AUG-1997; US-056881.
PR   22-AUG-1997; US-056882.
PR   22-AUG-1997; US-056884.
PR   22-AUG-1997; US-056886.
PR   22-AUG-1997; US-056887.
PR   22-AUG-1997; US-056888.
PR   22-AUG-1997; US-056889.
PR   22-AUG-1997; US-056892.
PR   22-AUG-1997; US-056893.
PR   22-AUG-1997; US-056894.
PR   22-AUG-1997; US-056903.
PR   22-AUG-1997; US-056908.
PR   22-AUG-1997; US-056909.
PR   22-AUG-1997; US-056910.
PR   22-AUG-1997; US-056911.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057669.
PR   05-SEP-1997; US-057761.
PR   12-SEP-1997; US-058785.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI   Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR   WPI; 98-506364/43.
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Disclosure; Page 43; 72pp; English.
CC   This sequence represents a fragment of a secreted human protein encoded
CC   by the nucleic acid molecule designated Gene 64 (V59574).
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
SQ   Sequence   31 AA;

Query Match   7.4%;   Score 111; DB 37;   Length 31;
Best Local Similarity 34.5%;   Pred. No. 7.85e-01;
Matches   10;   Conservative   8;   Mismatches 11;   Indels   0;   Gaps   0;

Db   3 rdyiceycarafksshnlavhrmihtgek 31
             | ||:| ::| :|:|:::|:| :| ::| :: |
Qy   64 RSTYCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT   5
ID   W75040 standard; Protein; 183 AA.
AC   W75040;
DT   25-JAN-1999 (first entry)
DE   Fragment of human secreted protein encoded by gene 64.
KW   Human; secreted protein; testis; tumour; foetal brain tissue;
KW   fusion protein; cancer; central nervous system; seizure;
KW   diagnosis; neurodegenerative disease.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   Misc-difference 58
FT   /label= unknown
FT   Misc-difference 78
FT   /label= unknown
FT   Misc-difference 98
FT   /label= unknown
FT   Misc-difference 150
FT   /label= unknown
FT   Misc-difference 183
FT   /label= unknown
PN   WO9839448-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; U04493.
PR   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.

```
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   23-MAY-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   13-JUN-1997; US-049610.
PR   08-JUL-1997; US-051926.
PR   16-JUL-1997; US-052874.
PR   18-AUG-1997; US-055724.
PR   22-AUG-1997; US-056630.
PR   22-AUG-1997; US-056631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-056862.
PR   22-AUG-1997; US-056864.
PR   22-AUG-1997; US-056874.
PR   22-AUG-1997; US-056875.
PR   22-AUG-1997; US-056876.
PR   22-AUG-1997; US-056877.
PR   22-AUG-1997; US-056878.
PR   22-AUG-1997; US-056879.
PR   22-AUG-1997; US-056880.
PR   22-AUG-1997; US-056881.
PR   22-AUG-1997; US-056882.
PR   22-AUG-1997; US-056884.
PR   22-AUG-1997; US-056886.
PR   22-AUG-1997; US-056887.
PR   22-AUG-1997; US-056888.
PR   22-AUG-1997; US-056889.
PR   22-AUG-1997; US-056892.
PR   22-AUG-1997; US-056893.
PR   22-AUG-1997; US-056894.
PR   22-AUG-1997; US-056903.

PR   22-AUG-1997; US-056908.
PR   22-AUG-1997; US-056909.
PR   22-AUG-1997; US-056910.
PR   22-AUG-1997; US-056911.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057669.
PR   05-SEP-1997; US-057761.
PR   12-SEP-1997; US-058785.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI   Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR   WPI: 98-506364/43.
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Disclosure: Page 43; 721pp; English.
CC   This sequence represents a fragment of a secreted human protein encoded
CC   by the nucleic acid molecule designated Gene 64 (V59574).
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
SQ   Sequence    183 AA;

Query Match         7.4%;  Score 111;  DB 37;  Length 183;
Best Local Similarity  34.5%;  Pred. No. 7.85e-01;
Matches  10;  Conservative   8;  Mismatches  11;  Indels   0;  Gaps   0;

Db        152 rdylceycarafksshnlavhrmihtgek 180
              | :|:| |  |:|||:|: |: | ::  ::
Qy         64 RSYTCSFCKREFRSAQALGGHMNVHRRDR 92


RESULT   6
ID   W74793 standard; Protein; 203 AA.
AC   W74793;
DT   25-JAN-1999  (first entry)
DE   Human secreted protein encoded by gene 64 clone HMWEX24.
KW   Human; secreted protein; testis; tumour; foetal brain tissue;
KW   fusion protein; cancer; central nervous system; seizure;
KW   diagnosis; neurodegenerative disease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc-difference 203
FT                   label= unknown

PN   WO983448-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; U04493.
PR   02-OCT-1997; US-061060.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   11-APR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
```

PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   11-APR-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047581.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   13-JUN-1997; US-049610.
PR   08-JUL-1997; US-051926.
PR   16-JUL-1997; US-052894.
PR   18-AUG-1997; US-055724.
PR   22-AUG-1997; US-056630.
PR   22-AUG-1997; US-056631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056662.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-056862.
PR   22-AUG-1997; US-056864.
PR   22-AUG-1997; US-056872.
PR   22-AUG-1997; US-056874.
PR   22-AUG-1997; US-056875.
PR   22-AUG-1997; US-056876.
PR   22-AUG-1997; US-056877.
PR   22-AUG-1997; US-056878.
PR   22-AUG-1997; US-056879.
PR   22-AUG-1997; US-056880.
PR   22-AUG-1997; US-056881.
PR   22-AUG-1997; US-056882.
PR   22-AUG-1997; US-056884.
PR   22-AUG-1997; US-056886.
PR   22-AUG-1997; US-056887.
PR   22-AUG-1997; US-056888.
PR   22-AUG-1997; US-056889.
PR   22-AUG-1997; US-056892.

PR   22-AUG-1997; US-056893.
PR   22-AUG-1997; US-056894.
PR   22-AUG-1997; US-056903.
PR   22-AUG-1997; US-056908.
PR   22-AUG-1997; US-056909.
PR   22-AUG-1997; US-056910.
PR   22-AUG-1997; US-056911.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057669.
PR   05-SEP-1997; US-057761.
PR   12-SEP-1997; US-058785.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI   Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR   WPI: 98-506364/43.
DR   N-PSDB; V59574.
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 1; Page 578; 721pp; English.
CC   This sequence represents a secreted human protein encoded by the nucleic
CC   acid molecule designated Gene 64 from the human cDNA clone HMWEX24
CC   (deposited as clone ATCC 97900 and ATCC 209046).
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. V59502) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
SQ   Sequence   203 AA;

Query Match   7.4%;   Score 111; DB 37; Length 203;
Best Local Similarity   34.5%;   Pred. No. 7.85e-01;
Matches   10; Conservative   8; Mismatches 11; Indels   0; Gaps   0;

Db   154 rdyiceycarafksshnlavhrmihtgek 182
           | | | :| | | | |::::| | :| | : | ::
Qy   64 RSYTCSFCKKREFRSAQALGGHMNVHRRDR 92

RESULT   7
ID   W01457 standard; Protein; 364 AA.
AC   W01457;
DT   22-FEB-1997 (first entry)
DE   E. coli O157:H7 F12 region expression product.
KW   Haemorrhagic colitis; haemolytic uraemic syndrome; O157:H7; probe;
KW   diagnosis; pathogen; immunogen.
OS   Escherichia coli strain O157:H7.
PN   WO9612405-A1.
PD   17-OCT-1996.
PF   12-APR-1996; US05150.
PR   14-APR-1995; US-423564.
PA   (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
PA   (UNIW) UNIV WASHINGTON.
PI   Bilge SS, Tarr PI, Vary JC;
DR   WPI: 96-477064/47.
DR   N-PSDB; T44510.
PT   Nucleic acid probe for detection of entero-haemorrhagic E. coli
PT   O157:H7 - useful for distinguishing it from entero-pathogenic E.
PT   coli O55:H7
PS   Disclosure; Page 16-17; 22pp; English.
CC   The expression products (W01456 and W01457) of the F12 region
CC   (T44510) of enterohaemorrhagic E. coli O157:H7 respectively show
CC   homology to the putative transmembrane export protein of Yersinia
CC   pseudotuberculosis and to the perosamine synthetase (rfbE gene

CC   product) of Vibrio cholerae. They are considered candidate
CC   immunogens for preparing antibody reagents for detection of the
CC   O157:H7 strain.
SQ   Sequence 364 AA;

Query Match   6.8%;   Score 101; DB 20;   Length 364;
Best Local Similarity 29.3%;   Pred. No. 4.70e+00;
Matches   24; Conservative 17; Mismatches 35; Indels 6; Gaps 6;

Db   283 slltrtaeereelrnhladklietrpvfypvhtmpmysekyqkhpiaedlgwrginlpsf 342
               :||| :| :|| : | :| | :||||| || : | || || :||
Qy   93 AIL-RQSPPRDINRYSLNLNLEPNPNFYPSHN-PSFSRKFPPFEMRK-LG-KGV-VPN- 146

Db   343 pslsneqviyicesinefysdk 364
               |::| | || ::| : |:|
Qy   147 NHLKSARGRFGVEKIDSFMQEK 168

RESULT 8
ID   W87557 standard; Protein; 88 AA.
AC   W87557;
DT   01-MAR-1999 (first entry)
DE   A nucleic acid binding protein.
KW   Nucleic acid binding protein; Cys2-His2 zinc finger class; diagnosis;
KW   genetic disorder; chimeric restriction enzyme; gene therapy.
OS   Synthetic.
PN   WO9853058-A1.
PD   26-NOV-1998.
PF   26-MAY-1998; G01512.
PR   23-MAY-1997; GB-010809.
PA   (MEDI-) MEDICAL RES COUNCIL.
PI   Choo Y, Isalan M, Klug A;
WPI; 99-024578/02.
DR   N-PSDB: V83634.
SQ   Sequence 88 AA;
PT   preparing Cys2-His2 zinc finger class nucleic acid binding proteins
PT   - capable of binding to a nucleic acid quadruplet, by mutating sites
PT   in model zinc finger domains according to a defined set of
PT   substitutions
PS   Disclosure: Page 44; 64pp; English.
CC   The present sequence represents a nucleic acid binding protein (NBP) of
CC   the Cys2-His2 zinc finger class. The NBP is capable of binding to a
CC   nucleic acid quadruplet in a target sequence. The specification
CC   describers a method of preparing the NBP. The method is used for
CC   designing nucleic acid binding proteins which are useful in medicine.
CC   The proteins are specifically engineered to recognise particular nucleic
CC   acid sequences and as such are suitable for diagnosis of genetic
CC   disorders. The proteins can be used in the manufacture of chimeric
CC   restriction enzymes, in which a nucleic acid cleaving domain is fused
CC   to a nucleic acid binding domain comprising a zinc finger. Fusion protein
CC   comprising NBP and an integrase, e.g. viral integrase, can be used to
CC   target nucleic acid sequences in vivo. In gene therapy applications, the
CC   method may be targeted to the delivery of functional genes into defective
CC   genes, or the delivery of nonsense nucleic acid in order to disrupt
CC   undesired nucleic acid. Genes may also be delivered to known, repetitive
CC   stretches of nucleic acid, e.g. centromeres, together with an activating
CC   sequence such as an LCR. NBP can be specifically used to knockout cells
CC   having mutant proteins e.g. mutant ras. They can also be used to modulate
CC   the action of transcription factors, e.g. the activity of HIV tat may be
CC   reduced by NBP specific for HIV TAR. NBP may also be coupled to toxic
CC   molecules, e.g. nucleases, which are capable of selectively destroying
CC   cells which comprise a mutation in their endogenous nucleic acid.
SQ   Sequence 88 AA;

Query Match   6.5%;   Score 97; DB 38;   Length 88;
Best Local Similarity 28.0%;   Pred. No. 9.48e+00;
Matches   14; Conservative 16; Mismatches 18; Indels 2; Gaps 2;

Db   40 mr-nfsrsdnltrhlrhtgekpfgqricmrnfrqadhlqehlkthtgek 88
              |:||: :|:| :::|::|||:|:::|:||:|||:|:| : | ::
Qy   44 LRDNYGNEDHLLGGLFSWPP-RSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 9

ID   W84299 standard; Protein; 88 AA.
AC   W84299;
DT   18-MAR-1999 (first entry)
DE   An anti-HIV zinc finger.
KW   Anti-HIV zinc finger; nucleic acid binding protein; NABP;
KW   Cys2-His2 zinc finger; detection; gene delivery; gene delivery; ds.
OS   Synthetic.
PN   WO9853059-A1.
PD   26-NOV-1998.
PF   26-MAY-1998; G01514.
PR   23-MAY-1997; GB-010807.
PA   (MEDI-) MEDICAL RES COUNCIL.
PI   Choo Y, Isalan M, Klug A;
WPI; 99-045308/04.
DR   N-PSDB: V99467.
PT   Preparation of nucleic acid binding proteins - by designing protein
PT   sequences of a Cys2-His2 zinc finger class based on a nucleic acid
PT   base triplet in a target nucleic acid sequence
PS   Example 5; Page 42; 62pp; English.
CC   The present sequence encodes an anti-HIV zinc finger. The zinc finger
CC   was made to exemplify the invention. THe specification describes a
CC   method for preparing a nucleic acid binding protein (NABP) of the
CC   Cys2-His2 zinc finger class capable of binding to a nucleic acid base
CC   triplet in a target nucleic acid sequence. Binding to the 5' base of
CC   the triplet by an alpha-helical zinc finger nucleic acid binding motif
CC   in the protein is determined as follows: (a) if the 5' base in the
CC   triplet is A, then position +6 in the alpha -helix is Glu, Asn or
CC   Val; (b) if the 5' base in the triplet is C, then position +6 in the
CC   alpha -helix is Ser, Thr, Val, Ala, Glu or Asn. The methods can be used
CC   for designing a protein which is capable of binding to any predefined
CC   nucleic acid sequence. The NABPs can be used for the detection of target
CC   nucleic acid molecules. They can also be used in gene therapy, e.g. for
CC   the delivery of functional genes into defective genes, or the delivery
CC   of nonsense nucleic acid to disrupt undesired nucleic acid.
SQ   Sequence 88 AA;

Query Match   6.5%;   Score 97; DB 39;   Length 88;
Best Local Similarity 28.0%;   Pred. No. 9.48e+00;
Matches   14; Conservative 16; Mismatches 18; Indels 2; Gaps 2;

Db   40 mr-nfsrsdnltrhlrhtgekpfgqricmrnfrqadhlqehlkthtgek 88
              |:||: :|:| :::|::|||:|:::|:||:|||:|:| : | ::
Qy   44 LRDNYGNEDHLLGGLFSWPP-RSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 10
ID   W23975 standard; Protein; 1060 AA.
AC   W23975;
DT   21-JUL-1998 (first entry)
DE   Homo sapiens 20q13 amplicon ZABC-1 predicted protein.
KW   20q13 amplicon; chromosome 20; tumour; detection; ZABC-1 gene;
KW   chromosomal abnormalities; probe; gene therapy; antisense inhibition;
KW   treatment; age-related macular degeneration; retinitis pigmentation;
KW   Leber's congenital amaurosis; zinc finger amplified in breast cancer.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   Misc_difference 163
FT   /note= "undefined"
FT   Misc_difference 164
FT   /note= "undefined"
FT   Misc_difference 165
FT   /note= "undefined"
FT   Misc_difference 166
FT   /note= "undefined"
PN   WO9802539-A1.
PD   22-JAN-1998.
PF   15-JUL-1997; U12343.
PR   17-JAN-1997; US-785532.
PR   15-JUL-1996; US-680395.
PR   16-OCT-1996; US-731499.
PA   (REGC) UNIV CALIFORNIA.
PI   Collins CC, Godfrey T, Gray JW, Hwang SI, Kowbel D,
PI   Rommens J;

```
DR   WPI; 98-110587/10.
DR   N-PSDB; V09023, V09024.
PT   New sequences from the 20q13 amplicon - used for detecting
PT   chromosomal abnormalities, particularly tumours, and for developing
PT   products for treating diseases
PS   Disclosure: Page 69; 91pp; English.
CC   The sequence is that of the predicted protein encoded by ZABC-1
CC   (zinc finger amplified in breast cancer). It's coding sequence maps
CC   to the core of the 20q13.2 amplicon and is overexpressed in primary
CC   tumours and breast cancer cell lines having 20q13.2 amplification.
CC   The sequence can be used as a probe for the detection of chromosomal
CC   abnormalities at 20q13. It and other sequences isolated from the 20q13
CC   amplicon are consistently amplified in primary tumours. These sequences
CC   are useful as probes or as probe targets for monitoring the relative
CC   copy number of corresponding sequences from a biological sample such as
CC   tumour cells. The sequences can also be used in therapeutic applications
CC   for modulating the expression of the endogenous gene or the activity
CC   of the gene product. Examples of therapeutic approaches include
CC   antisense inhibition of gene expression, gene therapy, and monoclonal
CC   antibodies that specifically bind the gene products. The products can
CC   also be used in the treatment of other diseases, e.g. age-related
CC   macular degeneration, Leber's congenital amaurosis and retinitis
CC   pigmentation.
SQ   Sequence   1060 AA;

Query Match            6.3%;     Score 94;    DB 30;    Length 1060;
Best Local Similarity   35.1%;   Pred. No. 1.59e+01;
Matches     13;   Conservative    8;   Mismatches 16;   Indels   0;   Gaps   0;

Db     376 csecgkafrtyhqlvlhsrvhkkdrragaesptmsvd 412
                  ||  |:|:||  |: ||    :||       :||
Qy      68 CSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDIN 104


RESULT   11
ID   R91305    standard; Protein; 423 AA.
AC   R91305;
DT   07-JUL-1996 (first entry)
DE   Transcription factor-IIIA.
KW   Human; transcription factor-IIIA; hTFIIIA; DNA binding protein;
KW   ribosome; zinc finger; diagnostic; probe; transcription control;
KW   antitumour; cancer; therapy.
OS   Homo sapiens.
PN   EP-704526-A1.
PD   03-APR-1996.
PF   05-SEP-1995; 113908.
PR   05-SEP-1994; JP-211022.
PA   (SAKA ) OTSUKA PHARM CO LTD.
PI   Fujiwara T, Ozaki K,  Shimada Y,  Shin S,  Takeda S;
DR   WPI; 96-173033/18.
DR   N-PSDB; T14037, T14038.
PT   Human Transcription Factor III A gene - useful in regulation of
PT   transcription and for diagnosis and treatment of e.g. cancer related
PT   diseases
PS   Claim 1: Page 8-10; 17pp; English.
CC   The sequence represents human transcription factor-IIIA (hTFIIIA),
CC   a DNA binding protein which is necessary for the initiation of 5S
CC   RNA gene transcription, binding to an internal control region of
CC   the 5S gene. The protein contains 9 zinc finger domains, which are
CC   homologous to the C2H2 finger domains of Xenopus TFIIIA, except for
CC   the 6th finger domain, which has only 3 amino acid residues between
CC   2 cysteine residues, instead of 5 amino acid residues for Xenopus
CC   TFIIIA. The protein (optionally in recombinant form) and encoding
CC   gene may be used in diagnosis, identification or therapy of
CC   hereditary diseases such as cancer, or other diseases resulting from
CC   abnormal transcriptional control, and to analyse the mechanisms
CC   involved in their activity.
SQ   Sequence   423 AA;

Query Match            6.2%;     Score 92;    DB 17;    Length 423;
Best Local Similarity   24.6%;   Pred. No. 2.24e+01;
Matches     15;   Conservative 16;   Mismatches 27;   Indels   3;   Gaps   3;
```

```
Db     277 cevcrktfkrkdylkqhmkthaperdvcr-cpregcgrtytv-fnlqshilsfheesrp 334
                  |-|::|: .:| :||::| .  | |::  :  ||::::  ::||::| .||  ::
Qy      68 CSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINR-YSLLNLNLEPNPNFYPSHNPS 126

Db     335 f 335
                -
Qy     127 F 127


RESULT   12
ID   R68743    standard; Protein; 706 AA.
AC   R68743;
DT   19-JUL-1995 (first entry)
DE   BCL-6 zinc finger protein.
KW   bcl-6 locus; non-Hodgkin lymphoma; B-cell lymphoma; B-lymphocyte;
KW   diagnostic; therapeutic; chromosome-3q27; translocation;
KW   proto-oncogene; diffuse large cell lymphoma; DLCL; zinc finger.
OS   Homo sapiens.
FH   Key           Location/Qualifiers
FT   region        520..541
FT                 /label= zinc-finger
FT   region        548..568
FT                 /label= zinc-finger
FT   region        576..596
FT                 /label= zing-finger
FT   region        604..624
FT                 /label= zinc-finger
FT   region        632..652
FT                 /label= zinc-finger
FT   region        660..681
FT                 /label= zinc-finger
PN   WO9429343-A.
PD   22-DEC-1994.
PF   09-JUN-1994; U06669.
PR   09-JUN-1993; US-074967.
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PA   (SLOK ) SLOAN KETTERING INST CANCER.
PI   Chaganti RSK,  Dalla-Favera R,  Chaganti RS;
DR   WPI; 95-036403/05.
DR   N-PSDB; Q68743.
PT   Nucleic acid from genetic locus bcl-6 - used to develop prods.
PT   for diagnosis and therapy of B-cell lymphoma and non-Hodgkin's
PT   lymphoma
PS   Disclosure; Page 90-94; 129pp; English.
CC   DNA was extd. from tumor tissue of 2 cases of IgM-producing diffuse-
CC   type B-cell NHL carrying the t(3;14)(q27;q32) translocation. DNA
CC   analysis showed that the breakpoints on 3q27 were located within 3
CC   kb of the same genomic locus, which was designated bcl-6. The
CC   human bcl-6 locus was identified by screening a phage cDNA library
CC   constructed from Bjab B-cell lymphoma mRNA. The zinc finger
CC   protein encoded by bcl-6 is given in R68743. BCL-6 is a
CC   proto-oncogene specifically involved in the pathogenesis of diffuse
CC   large cell lymphoma.
SQ   Sequence   706 AA;

Query Match            6.2%;     Score 92;    DB 12;    Length 706;
Best Local Similarity   30.0%;   Pred. No. 2.24e+01;
Matches     9;   Conservative 10;   Mismatches 11;   Indels   0;   Gaps   0;

Db     628 kpypceicgtrfrhiqtlkshlrihtgekp 657
                  :||:|| || |  ||| | ::|  ::
Qy      64 RSYTCSFCKREFRSAQALGGHMNVHRRDRA 93


RESULT   13
ID   W81756    standard; Protein; 803 AA.
AC   W81756;
DT   27-JAN-1999 (first entry)
DE   Myc-binding zinc-finger protein.
KW   Myc; zinc-finger protein; transcription modulation; gene therapy.
OS   Unknown.
PN   EP-875567-A2.
```

PD   04-NOV-1998.
PF   08-APR-1998; 106426.
PR   30-APR-1997; DE-018249.
PA   (BADI) BASF AG.
PI   Eilers M, Haenel F, Peukert K;
DR   WPI; 98-559433/48.
DR   N-PSDB; V64579.
PT   Myc-binding zinc finger protein - useful for identifying
PT   transcription modulating substances
PS   Claim 1; Page 10-12; 13pp; German.
CC   This sequence represents a novel Myc-binding zinc-finger protein which
CC   can be used in a method to identify transcription-modulating substances.
CC   This process involves incubating the protein with myc gene product under
CC   conditions such that a complex is formed between the two proteins then
CC   repeating this step in the presence of one or more test substances and
CC   determining the difference in protein complex formation between the first
CC   and second step. Substances for which the protein complex formation in
CC   the second step is different from that of the first are selected. The
CC   protein can be used to produce antibodies. The nucleic acid sequence can
CC   be used for gene therapy.
SQ   Sequence 803 AA;

Query Match   6.2%; Score 93; DB 37; Length 803;
Best Local Similarity   28.6%;
Matches   12; Conservative   11; Mismatches   19; Indels   0; Gaps   0;

Db   584 rphkcsvcskafvngdiskhiiihtgekpyicdkcgrgfnr 625
              | |::| ||:| | :|::|:||:| | :||
Qy   64 RSYTCFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINR 105

RESULT   14
ID   R95242 standard: Protein; 547 AA.
AC   R95242;
DT   13-AUG-1996 (first entry)
DE   HIC-1 polypeptide.
KW   HIC-1; hypermethylated in cancer; tumour suppressor gene; cancer;
KW   diagnosis; therapy; prognosis; zinc finger transcription factor.
OS   Homo sapiens.
PN   WO961487-Al.
PD   23-MAY-1996.
PF   15-NOV-1995; U14996.
PR   15-NOV-1994; US-340203.
PA   (UYJO) UNIV JOHNS HOPKINS SCHOOL MED.
PI   Baylin SB, Wales MM;
DR   WPI; 96-259576/26.
DR   N-PSDB; T29605.
PT   Polynucleotide encoding hyper:methylated in cancer (HIC)-7
PT   polypeptide - is tumour suppressor gene, useful for diagnosis,
PT   prognosis and treatment of cancer.
PS   Claim 1; Page 45-50; 71pp; English.
CC   HIC-1 polypeptide (R95242) is encoded by the HIC-1 gene (T29605), or
CC   hypermethylated in cancer, a novel tumour suppressor gene which is
CC   ubiquitously expressed in normal tissues, but underexpressed in tumour
CC   cells (e.g. breast, lung, colon, fibroblasts) where it is
CC   hypermethylated. The HIC-1 polypeptide is characterised as having
CC   a distinct amino acid homology to a highly conserved N-terminal
CC   motif, termed the 2in (zinc finger N-terminal) domain present in
CC   zinc finger transcription factors. It also has 5 Kruppel type
CC   Cys2-His2 zinc fingers. Detection of HIC-1 is used for the
CC   diagnosis or prognosis of cell proliferative diseases.
SQ   Sequence 547 AA;

Query Match   6.1%; Score 91; DB 17; Length 547;
Best Local Similarity   32.1%;
Matches   9; Conservative   6; Mismatches   13; Indels   0; Gaps   0;

Db   343 wltrpypctlcgkkftqrgtmtrhmrsh 370
              | ::||:||:| :: ||| |||
Qy   61 WPPRSYTCSFCKREFRSAQALGGHMNVH 88

RESULT   15

ID   W20841 standard: Protein; 957 AA.
AC   W20841;
DT   17-JUL-1997 (first entry)
DE   H. pylori transmembrane protein, 12ae10622orf16.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW   identification; binding compound; bacterium; life cycle; activator;
KW   bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW   diagnosis.
OS   Helicobacter pylori.
PN   WO964089-Al.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR) ASTRA AB.
PI   Berglindh OT, Smith D, Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T68094.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 73; Page 1249-51; 148lpp; English.
CC   This sequence represents a H. pylori protein likely to contain two
CC   membrane spanning regions.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence 957 AA;

Query Match   6.1%; Score 91; DB 22; Length 957;
Best Local Similarity   24.7%; Pred. No. 2.66e+01;
Matches   20; Conservative   22; Mismatches   36; Indels   3; Gaps   3;

Db   74 eskkefyafyqkhysfancdfakklnhllkshfkvenhcvrfenfinrlafymatgsgk 133
              |::.||.:| ||::|:|: .| .:: :| :| || :|:|
Qy   114 EPNPNFYPSHNPSFSRKFPFEMRKLGKGVVPNNHLKSARGRF-G-VEKIDSFMQEKECT 171

Db   134 tivilklvellsvamgmglip 154
              ||| | |:|:::|:|||:
Qy   172 TTVIKK-SEFLRLDLGIGLIS 191

Search completed: Wed Jul 14 17:26:55 1999
Job time : 19 secs.

```
  *******************************************
  *******************************************
  __   __  __   __   ___   __   __
 |  \/  | |  \ |  | |   | |  \ |  |
 |      | |   \|  | |___| |   \|  |  (TM)
 |__|\/|_| |__|\|__| |___| |__|\|__|
  *******************************************
  *******************************************

        Release 3.1A John F. Collins, Biocomputing Research Unit.
        Copyright (c) 1993-1998 University of Edinburgh, U.K.
               Distribution rights by Oxford Molecular Ltd

 *******************************************

 MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

 Run on:          Wed Jul 14 17:26:04 1999;  MasPar time 11.70 Seconds
                            705.211 Million cell updates/sec

 Tabular output not generated.

 Title:          >US-09-156-580-2
 Description:    (1-206) from US09156580A.pep
 Perfect Score:  1495
 Sequence:       1 MAGMDRNSFNSKYFKNKSIM..........IGLISESKEDDLELRLGST 206

 Scoring table:  PAM 150
                 Gap 11

 Searched:       122810 seqs, 40068593 residues

 Post-processing: Minimum Match 0%
                  Listing first 45 summaries

 Database:       pir60
                 1:pir1 2:pir2 3:pir3 4:pir4

 Statistics:     Mean 44.623;  Variance 88.995;  scale 0.501

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description	Pred. No.
1	364	24.3	204	2	S60325	transcription factor	1.01e-46
2	276	18.5	304	2	T02540	probable zinc finger	4.13e-30
3	263	17.6	180	2	D71448	probable Zn finger pr	1.01e-27
4	169	11.3	209	2	S55887	CCHH finger protein 7	2.60e-11
5	164	11.0	228	2	S55881	CCHH finger protein 1	1.69e-10
6	160	10.7	197	2	S55886	CCHH finger protein 6	7.49e-10
7	157	10.5	259	2	S55884	CCHH finger protein 4	2.26e-09
8	154	10.3	150	2	S55882	CCHH finger protein 3	6.80e-09
9	148	9.9	235	2	S55883	CCHH finger protein 8	5.98e-08
10	141	9.4	228	2	S55888	CCHH finger protein 5	7.23e-07
11	125	8.4	211	2	S55885	CCHH finger protein,	1.77e-04
12	119	8.0	273	2	T01985	zinc-finger protein,	1.29e-03
13	116	7.8	281	2	S19159	finger protein EPF1 -	3.40e-03
14	117	7.8	485	2	A40751	finger protein MZF1 -	4.46e-03
15	115	7.7	339	2	JC1442	transcription factor	4.87e-03
16	113	7.6	339	2	B34895	transcription factor	8.87e-03
17	114	7.6	354	2	I48722	zinc finger protein -	6.45e-03
18	113	7.6	365	2	A34895	5S RNA-binding protei	8.87e-03
19	114	7.6	455	2	I48724	zinc finger protein P	6.45e-03
20	113	7.6	594	2	JC5146	arylphorin gene speci	8.87e-03
21	112	7.5	273	2	S69193	probable finger prote	1.22e-02
22	110	7.4	308	2	A44496	transcription repress	2.28e-02
23	111	7.4	484	2	I39200	zinc finger protein C	1.67e-02
24		7.4	110	3005	2	S33642	homeotic protein zfh-	2.28e-02
25		7.3	109	281	2	I39199	C2H2-150 - human	3.11e-02
26		7.3	109	1350	2	S00647	finger protein - Afri	3.11e-02
27		7.2	107	261	2	S39045	probable finger prote	5.78e-02
28		7.2	107	286	2	T02641	probable C2H2 zinc fi	5.78e-02
29		7.2	107	289	2	A48913	KRAB zinc finger prot	5.78e-02
30		7.2	108	704	2	B53235	finger protein BRcore	4.24e-02
31		7.2	108	704	2	S21911	BRcore-NS-23 protein	4.24e-02
32		7.2	108	899	2	B48586	suppressor of hairy w	4.24e-02
33		7.2	107	1173	2	I50620	prockr2 - chicken (fr	5.78e-02
34		7.1	106	295	2	A46017	zinc finger protein z	7.85e-02
35		7.0	105	196	2	S05551	finger protein (clone	1.06e-01
36		7.0	104	264	2	A49124	zinc-finger/transcrip	1.44e-01
37		7.0	105	392	2	S11998	finger protein odd-sk	1.06e-01
38		7.0	104	423	2	A48152	zinc finger protein G	1.44e-01
39		7.0	104	577	2	S72227	finger protein sob -	1.44e-01
40		7.0	104	630	2	S37884	finger protein MSN4 -	1.44e-01
41		7.0	104	667	2	S63587	gene pacC protein - A	1.44e-01
42		7.0	104	678	2	S54308	DNA binding protein -	1.44e-01
43		7.0	104	744	2	S70619	finger protein bowel	1.44e-01
44		7.0	104	1355	2	S40022	spalt protein - fruit	1.44e-01
45		7.0	104	1402	2	S42748	finger protein - frui	1.44e-01

ALIGNMENTS

```
 RESULT          1
 ENTRY           S60325              #type complete
 TITLE           transcription factor SUPERMAN - Arabidopsis thaliana
 ALTERNATE_NAMES CCHH zinc finger protein
 ORGANISM        #formal_name Arabidopsis thaliana #common_name mouse-ear
                 cress
 DATE            06-Sep-1996 #sequence_revision 27-Feb-1997 #text_change
                 09-Sep-1997
 ACCESSIONS      S60325
 REFERENCE       S60325
 #authors        Sakai, H.; Medrano, L.J.; Meyerowitz, E.M.
 #journal        Nature (1995) 378:199-203
 #title          Role of SUPERMAN in maintaining Arabidopsis floral whorl
                 boundaries.
 #cross-references MUID:96069740
 #accession      S60325
 ##molecule_type mRNA
 ##residues      1-204 ##label SAK
 ##cross-references EMBL:U38946; NID:g1079668; PID:g1079669
 ##note          DNA sequencing was also done
 GENETICS
 #gene           SUP
 KEYWORDS        DNA binding; leucine zipper; nucleus; transcription factor;
                 transcription regulation; zinc finger
 SUMMARY         #length 204 #molecular-weight 23084 #checksum 5851

       Query Match        24.3%;      Score 364;  DB 2;  Length 204;
       Best Local Similarity 46.2%;   Pred. No. 1.01e-46;
       Matches   80; Conservative  34; Mismatches  38; Indels  21; Gaps  11;

 Db     33 DHDYLLG--FSWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSSTPSP 90
              :||| |||||||||||||||||||||||||||||||  | | |
 Qy     50 NEDHLLGGLFSWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAILR--QSP-------- 99

 Db     91 PYPNPNYSYSTMANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTK 150
              |:||||| : : :|: |: |::|: : :|:|:|:|
 Qy    100 P-RDIN-RYSLLNLNLEPN--P-NFYPSHNPSFSRKFPPFEMRKLG-KGV-VPNNHLKSA 152

 Db    151 KSSLLVEAGEATRFTSKD-ACKILRNDEIISLELEIGLINESEQDLDELRLG 202
              :| || ::: :| :||||||||||||||
 Qy    153 RGRFGVEKIDSFM-QEKECTTVIKKSEFLRLDLGIGLISESKEDDLELRLG 204

 RESULT          2
 ENTRY           T02540              #type complete
 TITLE           probable zinc finger protein - Arabidopsis thaliana
 ALTERNATE_NAMES F13M22.24 protein
```

ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE   05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
              05-Mar-1999
ACCESSIONS   T02540
REFERENCE   Z14198
#authors   Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby,
              M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage,
              A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
#submission   submitted to the EMBL Data Library, June 1998
#description Arabidopsis thaliana chromosome II BAC F13M22 genomic
              sequence.
#accession   T02540
##status   preliminary; translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues   1-304 ##label ROU
##cross-references EMBL:AC004684; NID:g3236234; PID:g3236256
GENETICS
#map_position 2
#note   F13M22.24
SUMMARY   #length 304 #molecular-weight 34125 #checksum 1987

Query Match   18.5%;   Score 276; DB 2; Length 304;
Best Local Similarity 86.8%; Pred. No. 4.13e-30;
Matches 33; Conservative 4; Mismatches 4; Indels 1; Gaps 0;

Db   34 WPPRSYTCSFCKREFRSAQALGGHMNVHRRDRARLKQA 71
         ||:|||||||||||||||||||||||||||||||| |:|
Qy   61 WPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT   3
ENTRY   D71448
TITLE   probable zn finger protein - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
#variety   columbia
DATE   03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
              05-Dec-1998
ACCESSIONS   D71448
              A71400
REFERENCE   Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
#authors   Dean, C.; Bergkamp, R.; Dirkse, W.; Van Staveren, M.;
              Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel,
              K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.;
              Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.;
              Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.;
              Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.;
              Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger,
              M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
              James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka,
              A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
              Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.;
              Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.;
              Rechman, S.; Ansorge, W.; Cooke, R.; Berger, C.; Delseny,
              M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
              Schueller, C.; Chalwatzis, N.
#journal   Nature (1998) 391:485-488
#title   Analysis of 1.9 Mb of contiguous sequence from chromosome 4
              of Arabidopsis thaliana.
#cross-references MUID:98121113
#accession   D71448
##status   preliminary; nucleic acid sequence not shown;
              translation not shown
##molecule_type DNA
##residues   1-180 ##label BEV
##cross-references GB:Z97344; NID:g2245126; PID:g2245140
GENETICS
#map_position 4COP9-4G3845
SUMMARY   #length 180 #molecular-weight 19788 #checksum 748

Query Match   17.6%;   Score 263; DB 2; Length 180;
Best Local Similarity 86.5%; Pred. No. 1.01e-27;

Matches 32; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db   34 TWPPRSYTCNFCRREFRSAQALGGHMNVHRRDRASSR 70
         :||||||||:||:|||||||||||||||||||||||
Qy   60 SWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAILR 96

RESULT   4
ENTRY   S55887
TITLE   CCHH finger protein 7 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE   28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
              09-Sep-1997
ACCESSIONS   S55887
REFERENCE   S55881
#authors   Tague, B.W.; Goodman, H.M.
#journal   Plant Mol. Biol. (1995) 28:267-279
#title   Characterization of a family of Arabidopsis zinc finger
              protein cDNAs.
#cross-references MUID:95322589
#accession   S55887
##molecule_type mRNA
##residues   1-209 ##label TAG
##cross-references GB:L39650; NID:g790684; PID:g790685
KEYWORDS   DNA binding; zinc finger
FEATURE
61-81   #region zinc finger CCHH motif
SUMMARY   #length 209 #molecular-weight 23166 #checksum 5844

Query Match   11.3%;   Score 169; DB 2; Length 209;
Best Local Similarity 47.2%;   Pred. No. 2.60e-11;
Matches 17; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

Db   56 PRVFSCNVCRRKFYSSQALGGHQNAHKRERTMAKRA 91
         || ::|::|:|::|||||||||| |:|:::: :::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT   5
ENTRY   S55881
TITLE   CCHH finger protein 1 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE   28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
              09-Sep-1997
ACCESSIONS   S55881; S71240
REFERENCE   S55881
#authors   Tague, B.W.; Goodman, H.M.
#journal   Plant Mol. Biol. (1995) 28:267-279
#title   Characterization of a family of Arabidopsis zinc finger
              protein cDNAs.
#cross-references MUID:95322589
#accession   S55881
##molecule_type mRNA
##residues   1-228 ##label TAG
##cross-references GB:L39644; NID:g790672; PID:g790673
REFERENCE   S71240
#authors   Goodman, H.M.; Gallant, P.; Keifer-Higgins, S.; Rubenfield,
              M.; Church, G.M.
#submission   submitted to the EMBL Data Library, April 1996
#description A 37.5 Kb sequence from Arabidopsis thaliana chromosome I.
#accession   S71240
##molecule_type DNA
##residues   1-228 ##label GOO
##cross-references EMBL:U53501; NID:g1297184; PID:g1297186
GENETICS
#map_position 1
KEYWORDS   DNA binding; zinc finger
FEATURE
70-90   #region zinc finger CCHH motif
SUMMARY   #length 228 #molecular-weight 25468 #checksum 8710

```
Query Match          11.0%;   Score 164;   DB 2;   Length 228;
Best Local Similarity 47.2%;  Pred. No. 1.69e-10;
Matches   17;  Conservative  12;  Mismatches  7;   Indels  0;   Gaps  0;

Db   65 PRVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRG 100
        ||  ||::|  |:|:|:||||| |-|:|::|:::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT   6
ENTRY         S55886      #type complete
TITLE         CCHH finger protein 6 - Arabidopsis thaliana
ORGANISM      #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE          28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
              09-Sep-1997
ACCESSIONS    S55886
REFERENCE     S55881
#authors      Tague, B.W.; Goodman, H.M.
#journal      Plant Mol. Biol. (1995) 28:267-279
#title        Characterization of a family of Arabidopsis zinc finger
              protein cDNAs.
#cross-references MUID:95322589
#accession    S55886
##molecule_type mRNA
##residues    1-197 ##label TAG
##cross-references GB:L39649; NID:g790682; PID:g790683
GENETICS
#map_position 1
KEYWORDS      DNA binding; zinc finger
FEATURE
   43-63        #region zinc finger CCHH motif
SUMMARY       #length 197 #molecular-weight 21414   #checksum 4370

Query Match          10.7%;   Score 160;   DB 2;   Length 197;
Best Local Similarity 47.2%;  Pred. No. 7.49e-10;
Matches   17;  Conservative  17;  Mismatches  11;  Indels  8;   Gaps  0;

Db   38 SRKYECQYCCREFANSQALGGHQNAHKKERQLLKRA 73
        |||:|  :| ::|||||| | :::| :|:::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT   7
ENTRY         S55884      #type complete
TITLE         CCHH finger protein 4 - Arabidopsis thaliana
ORGANISM      #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE          28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
              09-Sep-1997
ACCESSIONS    S55884
REFERENCE     S55881
#authors      Tague, B.W.; Goodman, H.M.
#journal      Plant Mol. Biol. (1995) 28:267-279
#title        Characterization of a family of Arabidopsis zinc finger
              protein cDNAs.
#cross-references MUID:95322589
#accession    S55884
##molecule_type mRNA
##residues    1-259 ##label TAG
##cross-references GB:L39647; NID:g790678; PID:g790679
GENETICS
#map_position 1
KEYWORDS      DNA binding; zinc finger
FEATURE
   87-107       #region zinc finger CCHH motif
SUMMARY       #length 259 #molecular-weight 28644   #checksum 3436

Query Match          10.5%;   Score 157;   DB 2;   Length 259;
Best Local Similarity 45.7%;  Pred. No. 2.26e-09;
Matches   16;  Conservative  12;  Mismatches  7;   Indels  0;   Gaps  0;

Db   83 RVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRA 117


Qy   64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT   8
ENTRY         S55882      #type complete
TITLE         CCHH finger protein 2 - Arabidopsis thaliana
ORGANISM      #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE          28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change
              09-Sep-1997
ACCESSIONS    S55882
REFERENCE     S55881
#authors      Tague, B.W.; Goodman, H.M.
#journal      Plant Mol. Biol. (1995) 28:267-279
#title        Characterization of a family of Arabidopsis zinc finger
              protein cDNAs.
#cross-references MUID:95322589
#accession    S55882
##molecule_type mRNA
##residues    1-150 ##label TAG
##cross-references GB:L39645; NID:g790674; PID:g790675
KEYWORDS      DNA binding; zinc finger
FEATURE
   54-74        #region zinc finger CCHH motif
SUMMARY       #length 150 #molecular-weight 16956   #checksum 9338

Query Match          10.3%;   Score 154;   DB 2;   Length 150;
Best Local Similarity 31.9%;  Pred. No. 6.80e-09;
Matches   23;  Conservative  22;  Mismatches  23;  Indels  4;   Gaps  3;

Db   49 PRVFSCNYCQRKFYSSQALGGHQNAHKLERTLAKKS--RELFRSSN-TVDSDQ-PYPFSG 104
        ||::|::||:|  |:|:|||||| | :::|  :|  : :|| :  :::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINRYSLLNLNLEPNPNFYPS 122

Db  105 RFELYGRGYQGF 116
        :| : :: ::
Qy  123 HNPSFSRKFPPF 134

RESULT   9
ENTRY         S55883      #type complete
TITLE         CCHH finger protein 3 - Arabidopsis thaliana
ORGANISM      #formal_name Arabidopsis thaliana #common_name mouse-ear
              cress
DATE          28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
              09-Sep-1997
ACCESSIONS    S55883
REFERENCE     S55881
#authors      Tague, B.W.; Goodman, H.M.
#journal      Plant Mol. Biol. (1995) 28:267-279
#title        Characterization of a family of Arabidopsis zinc finger
              protein cDNAs.
#cross-references MUID:95322589
#accession    S55883
##molecule_type mRNA
##residues    1-235 ##label TAG
##cross-references GB:L39646; NID:g790676; PID:g790677
GENETICS
#map_position 5
KEYWORDS      DNA binding; zinc finger
FEATURE
   63-83        #region zinc finger CCHH motif
SUMMARY       #length 235 #molecular-weight 26351   #checksum 6545

Query Match          9.9%;    Score 148;   DB 2;   Length 235;
Best Local Similarity 45.5%;  Pred. No. 5.98e-08;
Matches   15;  Conservative  12;  Mismatches  6;   Indels  0;   Gaps  0;

Db   61 FSCNYCQRTFYSSQALGGHQNAHKRERTLAKRG 93
        ::|:|:||| |:|:|||||| | :::| :|:::
Qy   66 YTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98
```

RESULT 10
ENTRY   S55888   #type complete
TITLE   CCHH finger protein 8 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE   28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
            25-Apr-1997
ACCESSIONS S55888
REFERENCE   S55881
#authors   Tague, B.W.; Goodman, H.M.
#journal   Plant Mol. Biol. (1995) 28:267-279
#title   Characterization of a family of Arabidopsis zinc finger
            protein cDNAs.
#cross-references MUID:95322589
#accession S55888
##molecule_type mRNA
##residues 1-228 ##label TAG
##cross-references GB:L39651
KEYWORDS   DNA binding; zinc finger
FEATURE
 84-104   #region zinc finger CCHH motif
SUMMARY   #length 228 #molecular-weight 26437 #checksum 4583

Query Match
Best Local Similarity   9.4%;   Score 141;   DB 2;   Length 228;
Matches   15; Conservative   7; Mismatches   7; Indels   0; Gaps   0;

Db   80 RRFECHYCFRNFPTSQALGGHQNAHKRER 108
         | : | : | : : | | | | | | : | : |
Qy   64 RSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 11
ENTRY   S55885   #type complete
TITLE   CCHH finger protein 5 - Arabidopsis thaliana
ORGANISM   #formal_name Arabidopsis thaliana #common_name mouse-ear
            cress
DATE   28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
            09-Sep-1997
ACCESSIONS S55885
REFERENCE   S55881
#authors   Tague, B.W.; Goodman, H.M.
#journal   Plant Mol. Biol. (1995) 28:267-279
#title   Characterization of a family of Arabidopsis zinc finger
            protein cDNAs.
#cross-references MUID:95322589
#accession S55885
##molecule_type mRNA
##residues 1-211 ##label TAG
##cross-references GB:L39648; NID:g790680; PID:g790681
GENETICS
#map_position 1
KEYWORDS   DNA binding; zinc finger
FEATURE
 62-82   #region zinc finger CCHH motif
SUMMARY   #length 211 #molecular-weight 23590 #checksum 5605

Query Match
Best Local Similarity   8.4%;   Score 125;   DB 2;   Length 211;
Matches   13; Conservative   8; Mismatches   8; Indels   0; Gaps   0;

Db   58 KRHECQYCGKEFANSQALGGHQNAHKKER 86
         | : | : : | | | | | | | : | : : |
Qy   64 RSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 12
ENTRY   T01985   #type complete
TITLE   zinc-finger protein, osmotic stress-induced - common tobacco
ORGANISM   #formal_name Nicotiana tabacum #common_name common tobacco
DATE   26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change
            26-Feb-1999

ACCESSIONS T01985
REFERENCE   Z14479
#authors   Reinhoud, P.J.; Versteege, I.G.A.H.; Logemann, S.; van Iren,
            F.; Kijne, J.W.
#submission submitted to the EMBL Data Library, March 1998
#description Mannitol preculture of tobacco suspension cells induces
            expression of a gene encoding a zinc-finger protein and of
            a lea5 gene.
#accession T01985
##status   preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-273 ##label REI
##cross-references EMBL:AF053077; NID:g2981168; PID:g2981169
##experimental_source cultivar White Burley
KEYWORDS   DNA binding; transcription factor; zinc finger
SUMMARY   #length 273 #molecular-weight 28892 #checksum 9922

Query Match
Best Local Similarity   8.0%;   Score 119;   DB 2;   Length 273;
Matches   15; Conservative   4; Mismatches   8; Indels   0; Gaps   0;

Db 103 QSYKCSVCDKAFSSYQALGGHKASHRK 129
         : : | | | | : | | : | | | | | | :
Qy   64 RSYTCSFCKREFRSAQALGGHMNVHRR 90

RESULT 13
ENTRY   S19159   #type complete
TITLE   finger protein EPF1 - garden petunia
ALTERNATE_NAMES DNA-binding protein
ORGANISM   #formal_name Petunia x hybrida #common_name garden petunia
DATE   13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
            09-Sep-1997
ACCESSIONS S19159
REFERENCE   S19159
#authors   Takatsuji, H.; Mori, M.; Benfey, P.N.; Ren, L.; Chua, N.H.
#journal   EMBO J. (1992) 11:241-249
#title   Characterization of a zinc finger DNA-binding protein
            expressed specifically in Petunia petals and seedlings.
#cross-references MUID:9215164
#accession S19159
##molecule_type mRNA
##residues 1-281 ##label TAK
##cross-references EMBL:X60700; NID:g20545; PID:g20546
GENETICS
#gene   EPF1
FUNCTION
#description may be one of the factors involved in the activation of
            5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)
KEYWORDS   DNA binding; transcription factor; zinc finger
FEATURE
 92-112   #region zinc finger CCHH motif\
 174-194   #region zinc finger CCHH motif
SUMMARY   #length 281 #molecular-weight 31321 #checksum 6401

Query Match
Best Local Similarity   7.8%;   Score 116;   DB 2;   Length 281;
Matches   14; Conservative   5; Mismatches 10; Indels   0; Gaps   0;

Db   90 YECKTCNRTFPSFQALGGHRTSHKKSKTI 118
         | : | | | | | : | | | | | : : : :
Qy   66 YTCSFCKREFRSAQALGGHMNVHRRDRAI 94

RESULT 14
ENTRY   A40751   #type complete
TITLE   finger protein MZF1 - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE   21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
            24-Sep-1998
ACCESSIONS A40751
REFERENCE   A40751
#authors   Hromas, R.; Collins, S.J.; Hickstein, D.; Raskind, W.;

```
#journal          Deaven, L.L.; O'Hara, P.; Hagen, F.S.; Kaushansky, K.
                  J. Biol. Chem. (1991) 266:14183-14187
#title            A retinoic acid-responsive human zinc finger gene, MZF-1,
                  preferentially expressed in myeloid cells.
#cross-references MUID:9131776i
#accession        A40751
  ##status        preliminary
  ##molecule_type mRNA
  ##residues      1-485 ##label HRO
  ##cross-references GB:M58297; NID:g189043; PID:g189044
GENETICS
  #gene           GDB:ZNF42; MZF-1
    ##cross-references GDB:125898; OMIM:194550
  #map_position   19q13.2-19q13.4
KEYWORDS          DNA binding; transcription regulation; zinc finger
SUMMARY           #length 485  #molecular-weight 54463  #checksum 2245


Query Match       7.8%;  Score 117;  DB 2;  Length 485;
Best Local Similarity  30.4%;  Pred. No. 2.46e-03;
Matches   14;  Conservative   14;  Mismatches  18;  Indels   0;  Gaps   0;

Db    430 RPYACPECGKAFRQRPTLTQHLRTHRREKPFACQDCGRRFHQSTKL 475
          |:|:|:| :| :||  :| |: |||:::|  | ::::| |
Qy     64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINRYSLL 109


RESULT  15
ENTRY             JC1442      #type complete
TITLE             transcription factor IIIA - American toad
ORGANISM          #formal_name Bufo americanus #common_name American toad
DATE              30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                  08-Sep-1997
ACCESSIONS        JC1442
REFERENCE         JC1441
  #authors        Gaskins, C.J.; Smith, J.F.; Ogilvie, M.K.; Hanas, J.S.
  #journal        Gene (1992) 120:197-206
  #title          Comparison of the sequence and structure of transcription
                  factor IIIA from Bufo americanus and Rana pipiens.
  #cross-references MUID:93013035
  #accession      JC1442
    ##molecule_type mRNA
    ##residues    1-339 ##label GAS
    ##cross-references EMBL:X58366; NID:g429028; PID:g429029
GENETICS
  #gene           TFIIIA
CLASSIFICATION    #superfamily transcription factor IIIA
KEYWORDS          DNA binding; duplication; nucleus; transcription factor; zinc
                  finger
SUMMARY           #length 339  #molecular-weight 39500  #checksum 9014


Query Match       7.7%;  Score 115;  DB 2;  Length 339;
Best Local Similarity  33.3%;  Pred. No. 4.69e-03;
Matches   17;  Conservative   15;  Mismatches  16;  Indels   3;  Gaps   3;

Db    191 TCSICNRTFKRKSFLKEHKKIHREERIVYR-CPRENCDRYTT-KFNLKSH 239
          |||:||:| |:  ||:::|| :| |   :| |  :| |: ::|| ::
Qy     67 TCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINR-YSLLNLNLEPN 116
```

Search completed: Wed Jul 14 17:26:18 1999
Job time : 14 secs.

THIS PAGE BLANK (USPTO)

```
*******************************************
*                                         *
*   MPSRCH                          (TM)  *
*                                         *
*******************************************
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jul 14 17:25:00 1999; MasPar time 8.03 Seconds
                                          725.174 Million cell updates/sec
Tabular output not generated.

Title: >US-09-156-580-2
Description: (1-206) from US09156580A.pep
Perfect Score: 1495
Sequence: 1 MAGMDRNSFNSKYFKNKSIM.........IGLISESKEDLDLELRLGST 206

Scoring table: PAM 150
               Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database: swiss-prot37
               1:swissprot

Statistics: Mean 45.702; Variance 77.977; scale 0.586

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES
```

| Result No. | Score | % Match | Query Length | DB | ID | Description | Pred. No. |
|---|---|---|---|---|---|---|---|
| 1 | 117 | 7.8 | 485 | 1 | ZN42_HUMAN | ZINC FINGER PROTEIN 42 | 2.27e-04 |
| 2 | 115 | 7.7 | 339 | 1 | TF3A_BUFAM | TRANSCRIPTION FACTOR I | 4.74e-04 |
| 3 | 115 | 7.7 | 397 | 1 | YKQ8_CAEEL | HYPOTHETICAL ZINC FING | 4.74e-04 |
| 4 | 113 | 7.6 | 365 | 1 | P43_XENBO | P43 5S RNA BINDING PRO | 9.84e-04 |
| 5 | 110 | 7.4 | 308 | 1 | AEF1_DROME | ADULT ENHANCER FACTOR | 2.90e-03 |
| 6 | 111 | 7.4 | 512 | 1 | Z169_HUMAN | ZINC FINGER PROTEIN 16 | 2.03e-03 |
| 7 | 110 | 7.4 | 3005 | 1 | ZFH2_DROME | ZINC-FINGER PROTEIN 2 | 2.90e-03 |
| 8 | 109 | 7.3 | 1350 | 1 | XFIN_XENLA | XFIN PROTEIN. | 4.14e-03 |
| 9 | 107 | 7.2 | 289 | 1 | ZN75_HUMAN | ZINC FINGER PROTEIN 75 | 8.39e-03 |
| 10 | 108 | 7.2 | 704 | 1 | BRC3_DROME | BROAD-COMPLEX CORE-NS- | 5.90e-03 |
| 11 | 108 | 7.2 | 899 | 1 | SUHW_DROVI | SUPPRESSOR OF HAIRY WI | 5.90e-03 |
| 12 | 105 | 7.0 | 196 | 1 | ZG8_XENLA | GASTRULA ZINC FINGER P | 1.69e-02 |
| 13 | 104 | 7.0 | 264 | 1 | SNAI_MOUSE | SNAIL PROTEIN HOMOLOG | 2.38e-02 |
| 14 | 105 | 7.0 | 392 | 1 | ODD_DROME | ODD-SKIPPED PROTEIN. | 1.69e-02 |
| 15 | 104 | 7.0 | 423 | 1 | ZNG1_RAT | ZINC FINGER PROTEIN GF | 2.38e-02 |
| 16 | 104 | 7.0 | 630 | 1 | MSN4_YEAST | ZINC FINGER PROTEIN MS | 2.38e-02 |
| 17 | 104 | 7.0 | 1355 | 1 | SALM_DROVI | HOMEOTIC PROTEIN SPALT | 2.38e-02 |
| 18 | 104 | 7.0 | 1402 | 1 | SALM_DROVI | HOMEOTIC PROTEIN SPALT | 2.38e-02 |
| 19 | 103 | 6.9 | 1191 | 1 | ZN91_HUMAN | ZINC FINGER PROTEIN 91 | 3.36e-02 |
| 20 | 101 | 6.8 | 604 | 1 | GLAS_DROME | GLASS PROTEIN. | 6.63e-02 |
| 21 | 102 | 6.8 | 625 | 1 | R101_YEAST | REGULATORY PROTEIN RIM | 4.73e-02 |
| 22 | 101 | 6.8 | 636 | 1 | ZF90_MOUSE | ZINC FINGER PROTEIN 90 | 6.63e-02 |
| 23 | 102 | 6.8 | 728 | 1 | Y413_MYCGE | HYPOTHETICAL PROTEIN M | 4.73e-02 |
| 24 | 102 | 6.8 | 1036 | 1 | Y414_MYCGE | HYPOTHETICAL PROTEIN M | 4.73e-02 |
| 25 | 100 | 6.7 | 336 | 1 | ZG57_XENLA | GASTRULA ZINC FINGER P | 9.29e-02 |
| 26 | 100 | 6.7 | 339 | 1 | TF3A_XENBO | TRANSCRIPTION FACTOR I | 9.29e-02 |
| 27 | 100 | 6.7 | 474 | 1 | Z141_HUMAN | ZINC FINGER PROTEIN 14 | 9.29e-02 |
| 28 | 100 | 6.7 | 510 | 1 | CF2_DROME | CHORION TRANSCRIPTION | 9.29e-02 |
| 29 | 100 | 6.7 | 1195 | 1 | YK76_YEAST | HYPOTHETICAL 137.5 KD | 9.29e-02 |
| 30 | 99 | 6.6 | 589 | 1 | Z132_HUMAN | ZINC FINGER PROTEIN 13 | 1.30e-01 |
| 31 | 99 | 6.6 | 678 | 1 | YNC7_YEAST | HYPOTHETICAL 76.3 KD Z | 1.30e-01 |
| 32 | 98 | 6.6 | 886 | 1 | SUHW_DROAN | SUPPRESSOR OF HAIRY WI | 1.81e-01 |
| 33 | 99 | 6.6 | 1687 | 1 | Z142_HUMAN | ZINC FINGER PROTEIN 14 | 1.30e-01 |
| 34 | 97 | 6.5 | 79 | 1 | KRUH_DROME | KRUEPPEL HOMOLOGOUS PR | 2.51e-01 |
| 35 | 97 | 6.5 | 466 | 1 | KRUP_DROME | KRUEPPEL PROTEIN. | 2.51e-01 |
| 36 | 96 | 6.4 | 259 | 1 | SNAI_XENLA | SNAIL PROTEIN HOMOLOG. | 3.48e-01 |
| 37 | 96 | 6.4 | 407 | 1 | OZF_MOUSE | ZINC FINGER PROTEIN OZ | 3.48e-01 |
| 38 | 96 | 6.4 | 511 | 1 | EGR1_BRARE | EARLY GROWTH RESPONSE | 3.48e-01 |
| 39 | 96 | 6.4 | 644 | 1 | BTD_DROME | TRANSCRIPTION FACTOR B | 3.48e-01 |
| 40 | 95 | 6.4 | 654 | 1 | Z133_HUMAN | ZINC FINGER PROTEIN 13 | 4.82e-01 |
| 41 | 96 | 6.4 | 675 | 1 | ZG20_XENLA | GASTRULA ZINC FINGER P | 3.48e-01 |
| 42 | 95 | 6.4 | 704 | 1 | MSN2_YEAST | ZINC FINGER PROTEIN MS | 4.82e-01 |
| 43 | 94 | 6.3 | 420 | 1 | ZG58_XENLA | GASTRULA ZINC FINGER P | 6.64e-01 |
| 44 | 94 | 6.3 | 1380 | 1 | ZMS1_YEAST | ZINC FINGER PROTEIN ZM | 6.64e-01 |
| 45 | 93 | 6.2 | 474 | 1 | MIG1_KLULA | REGULATORY PROTEIN MIG | 9.14e-01 |

```
ALIGNMENTS

RESULT  1
ID   ZN42_HUMAN     STANDARD;      PRT;     485 AA.
AC   P28698;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN 42 (MYELOID ZINC FINGER 1) (MZF-1).
GN   ZNF42.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91317761.
RA   HROMAS R., COLLINS S.J., HICKSTEIN D., RASKIND W., DEAVEN L.L.,
RA   O'HARA P., HAGEN F.S., KAUSHANSKY K.;
RT   "A retinoic acid-responsive human zinc finger gene, MZF-1,
RT   preferentially expressed in myeloid cells.";
RL   J. BIOL. CHEM. 266:14183-14187(1991).
CC   -!- FUNCTION: MAY BE ONE REGULATOR OF TRANSCRIPTIONAL EVENTS DURING
CC       HEMOPOIETIC DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN DIFFERENTIATING
CC       MYELOID CELLS.
CC   -!- INDUCTION: BY RETINOIC ACID.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M58297; G189044; -.
DR   PIR; A40751; A40751.
DR   MIM; 194550; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 13.
DR   PFAM; PF00096; zf-C2H2; 13.
DR   HSSP; P25490; 12NM.
DR   TRANSFAC; T00529; -.
KW   TRANSCRIPTION REGULATION; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
FT   DOMAIN       61     72       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      107    213       4 C2H2-TYPE ZINC-FINGERS.
FT   DOMAIN      214    235       GLY/PRO-RICH.
```

```
FT   DOMAIN        236    482    9 C2H2-TYPE ZINC-FINGERS.
FT   ZN_FING       107    129    C2H2-TYPE.
FT   ZN_FING       135    157    C2H2-TYPE.
FT   ZN_FING       163    185    C2H2-TYPE.
FT   ZN_FING       191    213    C2H2-TYPE.
FT   ZN_FING       236    258    C2H2-TYPE.
FT   ZN_FING       264    286    C2H2-TYPE.
FT   ZN_FING       292    314    C2H2-TYPE.
FT   ZN_FING       320    342    C2H2-TYPE.
FT   ZN_FING       348    370    C2H2-TYPE.
FT   ZN_FING       376    398    C2H2-TYPE.
FT   ZN_FING       404    426    C2H2-TYPE.
FT   ZN_FING       432    454    C2H2-TYPE.
FT   ZN_FING       460    482    C2H2-TYPE.
SQ   SEQUENCE      485 AA;  54463 MW;  908B33F1 CRC32;

Query Match             7.8%;
Best Local Similarity  30.4%;   Score 117;  DB 1;  Length 485;
                                Pred. No. 2.27e-04;
Matches   14;  Conservative  14;  Mismatches  18;  Indels  0;  Gaps  0;

Db   430  RPYACPECGKAFQRPTLTQHLRTHRREKPFACQDCGRRFHQSTKL 475
          :|:|::|| : | ||::||:::| |:|::|
Qy    64  RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINRYSLL 109


RESULT 2
ID   TF3A_BUFAM   STANDARD;   PRT;   339 AA.
AC   P34694;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE   TRANSCRIPTION FACTOR IIIA (FACTOR A) (TFIIIA).
OS   BUFO AMERICANUS (AMERICAN TOAD).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   NEOBATRACHIA; BUFONOIDEA; BUFONIDAE; BUFO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RX   MEDLINE; 93013035.
RA   GASKINS C.J., SMITH J.F., OGILVIE M.K., HANAS J.S.;
RT   "Comparison of the sequence and structure of transcription factor
RT   IIIA from Bufo americanus and Rana pipiens.";
RL   GENE 120:197-206(1992).
CC   -!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC       APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC       CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC       BINDS THE TRANSCRIBED 5S RNA'S.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; X58366; G429029; -.
DR   PIR; JC1442; JC1442.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
DR   PFAM; PF00096; zf-C2H2; 9.
DR   HSSP; P03001; 1TF6.
KW   TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
KW   RNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT   DOMAIN         12    274    ZINC-FINGERS.
FT   ZN_FING        12     36    C2H2-TYPE.
FT   ZN_FING        42     66    C2H2-TYPE.
FT   ZN_FING        72     97    C2H2-TYPE.
FT   ZN_FING       104    128    C2H2-TYPE.
FT   ZN_FING       134    158    C2H2-TYPE.
FT   ZN_FING       161    187    C2H2-TYPE.
FT   ZN_FING       191    212    C2H2-TYPE.
FT   ZN_FING       219    244    C2H2-TYPE.
```

```
FT   ZN_FING       250    274    C2H2-TYPE.
SQ   SEQUENCE      339 AA;  39500 MW;  AD3BD4A3 CRC32;

Query Match             7.8%;
Best Local Similarity  33.3%;   Score 115;  DB 1;  Length 339;
                                Pred. No. 4.74e-04;
Matches   17;  Conservative  15;  Mismatches  16;  Indels  3;  Gaps  3;

Db   191  TCSICNRTFKRKSFLKEHKKIHRERIVYR-CPRNCDRTYTT-KFNLKSH 239
          |:||::|:|: ::||: | :|:::||::| |::: ||: ::||
Qy    67  TCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINR-YSLLNLNLEPN 116


RESULT 3
ID   YKQ8_CAEEL   STANDARD;   PRT;   397 AA.
AC   P34303;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL ZINC FINGER PROTEIN C06E1.8 IN CHROMOSOME III.
GN   C06E1.8.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC   RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA   JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
RA   LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   NATURE 368:32-38(1994).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: WITH OTHER ZINC-FINGER PROTEINS.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the European outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; L16559; G289625; -.
DR   WORMPEP; C06E1.8; CE00063.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
DR   PFAM; PF00096; zf-C2H2; 3.
DR   HSSP; P25490; 1UBD.
KW   HYPOTHETICAL PROTEIN; ZINC-FINGER; DNA-BINDING; METAL-BINDING;
KW   NUCLEAR PROTEIN.
FT   ZN_FING        51     74    C2H2-TYPE.
FT   ZN_FING        78    102    C2H2-TYPE.
FT   ZN_FING       107    125    C2H2-TYPE.
FT   DOMAIN        131    140    ASN-RICH.
SQ   SEQUENCE      397 AA;  44476 MW;  FD85FAD4 CRC32;

Query Match             7.8%;
Best Local Similarity  27.4%;   Score 115;  DB 1;  Length 397;
                                Pred. No. 4.74e-04;
Matches   17;  Conservative  17;  Mismatches  24;  Indels  1;  Gaps  1;

Db    76  KSYKCTVCRKDISSSESLRTHMFKQHHISRMYMCRCCNWAFPDKSLLHIHLQTATNNNDS 135
          |::|::|: : |:::| :|:::| |||::| :: :|| : |:::|||:|||::
Qy    64  RSYTCSFCKREFRSAQALGGHM-NVHRRDRAILRQSPPRDINRYSLLNLNLEPNPNFYPS 122
```

Db   136 NN 137
          :|
Qy   123 HN 124

RESULT 4
ID   P43_XENBO   STANDARD;   PRT;   365 AA.
AC   P25066;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   P43 5S RNA BINDING PROTEIN (42S P43) (THESAURIN B).
OS   XENOPUS BOREALIS (KENYAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90235278.
RA   JOHO K.E., DARBY M.K., CRAWFORD E.T., BROWN D.D.;
RT   "A finger protein structurally similar to TFIIIA that binds
RT   exclusively to 5S RNA in Xenopus.";
RL   CELL 61:293-300(1990).
CC   -!- FUNCTION: P43 IS A 5S RNA BINDING PROTEIN WHICH IS A MAJOR
CC   CONSTITUENT OF OOCYTES AND COMPRISES PART OF A 42S
CC   RIBONUCLEOPROTEIN STORAGE PARTICLE.
CC   -!- SUBUNIT: THE 42S RNP PARTICLE COMPRISES FOUR SUBUNITS EACH OF
CC   WHICH CONTAINS ONE MOLECULE OF 5S RNA, THREE MOLECULES OF TRNA,
CC   TWO MOLECULES OF P50 (EF1-ALPHA) AND ONE MOLECULE OF THE 5S RNA
CC   BINDING PROTEIN 43.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M32471; G214149; -.
DR   PIR; A34895; A34895.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 7.
DR   PFAM; PF00096; zf-C2H2; 9.
DR   HSSP; P25490; 1UBD.
KW   ZINC-FINGER; METAL-BINDING; RNA-BINDING; REPEAT.
FT   DOMAIN   15   275   ZINC-FINGERS.
FT   ZN_FING   15   39   C2H2-TYPE.
FT   ZN_FING   45   69   C2H2-TYPE.
FT   ZN_FING   75   100   C2H2-TYPE.
FT   ZN_FING   106   130   C2H2-TYPE.
FT   ZN_FING   136   160   C2H2-TYPE.
FT   ZN_FING   163   187   C2H2-TYPE.
FT   ZN_FING   191   213   C2H2-TYPE.
FT   ZN_FING   220   245   C2H2-TYPE.
FT   ZN_FING   251   275   C2H2-TYPE.
SQ   SEQUENCE   365 AA; 41726 MW; 69FE1B78 CRC32;

Query Match 7.6%; Score 113; DB 1; Length 365;
Best Local Similarity 24.8%; Pred. No. 9.84e-04;
Matches 35; Conservative 34; Mismatches 61; Indels 11; Gaps 11;

Db 193 CAACKKPFKKASALRRHKATHAKNPLQL-PCPRQDCDKIFSTVFNLTHHLRKVHLCLQTH 251
         :|:||:||:|||: :|: :||| :| ::|| :| :|:||
Qy   68 CSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINR-YS-LLNLNLE-PNPNFY-PSH 123

Db 252 RCPHSNCTRSFAMRESLVRHLVVHDPERKLKLKFG-RRPSKFLGRGTRCPTPVVEE-DL 309
         :| :: || : || : :| :|| :: : :|::
Qy 124 NPSFSRKFPPFEMRK-LGKG-VVPNNHLKSARGRFGVEKIDSFM-QEKECTTTVIKKSEF 180

Db 310 SHL-FSRKLLFHYKTRLETNL 329
         |: |:| :| :|:|
Qy 181 LRLDLGLGLISESKEDLDLEL 201

RESULT 5
ID   AEF1_DROME   STANDARD;   PRT;   308 AA.
AC   P39413;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   ADULT ENHANCER FACTOR 1 (AEF-1).
GN   AEF1.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OVARY;
RX   MEDLINE; 92375078.
RA   FALB D., MANIATIS T.;
RT   "Drosophila transcriptional repressor protein that binds specifically
RT   to negative control elements in fat body enhancers.";
RL   MOL. CELL. BIOL. 12:4093-4103(1992).
CC   -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR THAT BINDS SPECIFICALLY TO FAT
CC   BODY-SPECIFIC ENHANCERS, NAMELY THE ADULT ADH ENHANCER (AAE) AND
CC   THE ENHANCER THAT CONTROLS YOLK PROTEIN GENE EXPRESSION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: FOUND IN ALL TISSUES EXAMINED INCLUDING THE
CC   OVARY AND THE FAT BODY.
CC   -!- DEVELOPMENTAL STAGE: HIGHER LEVELS ARE FOUND IN THIRD-INSTAR
CC   LARVAE AND IN ADULTS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M90755; -; NOT_ANNOTATED_CDS.
DR   PIR; A44496; A44496.
DR   FLYBASE; FBgn0005594; Aef1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   HSSP; P25490; 1ZNM.
KW   TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING; NUCLEAR PROTEIN;
KW   ZINC-FINGER.
FT   DOMAIN   56   76   GLN-RICH (OPA-REPEAT).
FT   DOMAIN   71   76   POLY-GLN.
FT   DOMAIN   105   127   ALA-RICH.
FT   DOMAIN   132   135   POLY-PRO.
FT   ZN_FING   184   206   C2H2-TYPE.
FT   ZN_FING   212   234   C2H2-TYPE.
FT   ZN_FING   240   262   C2H2-TYPE.
FT   ZN_FING   268   290   C2H2-TYPE.
SQ   SEQUENCE   308 AA; 33782 MW; 3B1BD499 CRC32;

Query Match 7.4%; Score 110; DB 1; Length 308;
Best Local Similarity 29.0%; Pred. No. 2.90e-03;
Matches 9; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 238 KPYNCNFCPKHFRQLSTLANHVKIHTGEKPF 268
         :|| :|| |: |:|| ::|:
Qy   64 RSTTCSFCKREFRSAQALGGHMNVHRRDRAI 94

RESULT 6
ID   Z169_HUMAN   STANDARD;   PRT;   512 AA.
AC   Q14929;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN 169 (FRAGMENT).
GN   ZNF169.
OS   HOMO SAPIENS (HUMAN).

```
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97225201.
RA   CHIDAMBARAM A., GAILANI M., GERRARD B., STEWART C., GOLDSTEIN A.,
RA   CHUMAKOV I., BALE A.E., DEAN M.;
RT   "Characterization of a YAC contig containing the NBCCS locus and a
RT   novel Kruppel-type zinc finger sequence on chromosome segment
RT   9q22.3.";
RL   GENES CHROMOSOMES CANCER 18:212-218(1997).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN KIDNEY, IS WEAKLY
CC       EXPRESSED IN HEART, LIVER, SPLEEN, AND SMALL INTESTINE, AND IS NOT
CC       EXPRESSED IN ADULT BRAIN OR SPINAL CORD.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U28251; G903599; -.
DR   MIM; 603404; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 9.
DR   PFAM; PF00096; zf-C2H2; 11.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING; NUCLEAR PROTEIN.
FT   NON_TER        1      1
FT   DOMAIN       144    473       ZINC-FINGERS.
FT   ZN_FING      144    166       C2H2-TYPE.
FT   ZN_FING      172    194       C2H2-TYPE.
FT   ZN_FING      200    222       C2H2-TYPE.
FT   ZN_FING      228    250       C2H2-TYPE.
FT   ZN_FING      311    333       C2H2-TYPE.
FT   ZN_FING      339    361       C2H2-TYPE.
FT   ZN_FING      367    389       C2H2-TYPE.
FT   ZN_FING      395    417       C2H2-TYPE.
FT   ZN_FING      449    473       C2H2-TYPE.
FT   ZN_FING      479    504       C2H2-TYPE.
SQ   SEQUENCE     512 AA;   57651 MW;  0FA6FE76 CRC32;

Query Match        7.4%;  Score 111;  DB 1;  Length 512;
Best Local Similarity  29.8%;  Pred. No. 2.03e-03;
Matches 14;  Conservative 13;  Mismatches 19;  Indels 1;  Gaps 1;

Db  198 KPYVCRECGRHFRYTSSLTNHKRIHSGERPFVCQECGRGFRQKIALL 244
         :|| :| :|| |  :| :| :|| :|  :||::|| :  :|| : :||
Qy   64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDI-NRYSLL 109

RESULT  7
ID   ZFH2_DROME     STANDARD;      PRT;   3005 AA.
AC   P28167;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN   ZFH-2.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92001539.
RA   FORTINI M.E., LAI Z., RUBIN G.M.;
RT   "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
RT   containing both zinc-finger and homeodomain motifs.";
RL   MECH. DEV. 34:113-122(1991).

CC   -!- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC       NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.
CC   -!- SIMILARITY: CONTAINS THREE HOMEOBOX DOMAINS.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; M63450; G158823; -.
DR   PIR; S27817; S27817.
DR   PIR; S33642; S33642.
DR   FLYBASE; FBgn0004607; zfh2.
DR   PROSITE; PS00027; HOMEOBOX_1; 2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
DR   PROSITE; PS50071; HOMEOBOX_2; 3.
DR   PFAM; PF00046; homeobox; 3.
DR   PFAM; PF00096; zf-C2H2; 12.
DR   HSSP; P15822; 42NF.
DR   TRANSFAC; T00920; -.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING; HOMEOBOX; NUCLEAR PROTEIN;
KW   REPEAT.
FT   ZN_FING      133    156       C2H2-TYPE.
FT   ZN_FING      559    582       C2H2-TYPE.
FT   ZN_FING      614    638       C2H2-TYPE.
FT   ZN_FING      732    756       C2H2-TYPE.
FT   ZN_FING      897    916       C2H2-TYPE (DEGENERATE).
FT   ZN_FING      940    964       C2H2-TYPE.
FT   ZN_FING      999   1023       C2H2-TYPE.
FT   ZN_FING     1074   1098       C2H2-TYPE.
FT   ZN_FING     1210   1233       C2H2-TYPE.
FT   ZN_FING     1341   1365       C2H2-TYPE.
FT   ZN_FING     1438   1462       C2H2-TYPE.
FT   ZN_FING     1477   1500       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     1513   1535       C2H2-TYPE.
FT   ZN_FING     1541   1564       C2H2-TYPE.
FT   DNA_BIND    1797   1856       HOMEOBOX 1.
FT   DNA_BIND    2154   2213       HOMEOBOX 2.
FT   ZN_FING     2234   2256       C2H2-TYPE.
FT   ZN_FING     2371   2393       C2H2-TYPE.
FT   DNA_BIND    2760   2819       HOMEOBOX 3.
SQ   SEQUENCE    3005 AA;  332056 MW;  8B4CC45F CRC32;

Query Match        7.4%;  Score 110;  DB 1;  Length 3005;
Best Local Similarity  68.4%;  Pred. No. 2.90e-03;
Matches 13;  Conservative 3;  Mismatches 3;  Indels 0;  Gaps 0;

Db 1543 CSFCQRNFRSTQALQKHME 1561
         |||| |:|||:|||| ||:
Qy   68 CSFCKREFRSAQALGGHMN 86

RESULT  8
ID   XFIN_XENLA     STANDARD;      PRT;   1350 AA.
AC   P08045;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   XFIN PROTEIN.
GN   XFIN.
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 88082679.
RA   RUIZ I ALTABA R., PERRY-O'KEEFE H., MELTON D.A.;
```

```
RT   "Xfin: an embryonic gene encoding a multifingered protein in
RT   Xenopus.";
RL   EMBO J. 6:3065-3070(1987).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE; 94021366.
RA   ANDREAZZOLI M., DE LUCCHINI S., COSTA M., BARSACCHI G.;
RT   "RNA binding properties and evolutionary conservation of the Xenopus
RT   multifinger protein Xfin.";
RL   NUCLEIC ACIDS RES. 21:4218-4225(1993).
RN   [3]
RP   STRUCTURE BY NMR OF FINGER 31.
RX   MEDLINE; 89346749.
RA   LEE M.S., GIPPERT G.P., SOMAN K.V., CASE D.A., WRIGHT P.E.;
RT   "Three-dimensional solution structure of a single zinc finger DNA-
RT   binding domain.";
RL   SCIENCE 245:635-637(1989).
RN   [4]
RP   STRUCTURE BY NMR OF A FINGER.
RX   MEDLINE; 89378224.
RA   LEE M.S., CAVANAGH J., WRIGHT P.E.;
RT   "Complete assignment of the 1H NMR spectrum of a synthetic zinc
RT   finger from Xfin. Sequential resonance assignments and secondary
RT   structure.";
RL   FEBS LETT. 254:159-164(1989).
CC   -!- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
CC       REGULATION PROCESSES.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
CC       AS NEURAL RETINA CONES.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
CC       EMBRYOGENESIS.
CC   -!- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
CC   -!- PTM: PHOSPHORYLATED.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 KRAB BOX.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; X06021; G655234; -.
DR   PIR; S00647, S00647.
DR   PDB; 1ZNF; 15-OCT-91.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 35.
DR   PFAM; PF00096; zf-C2H2; 36.
KW   ZINC-FINGER; METAL-BINDING; RNA-BINDING; REPEAT; 3D-STRUCTURE;
KW   PHOSPHORYLATION.
FT   DOMAIN        1     26       KRAB BOX ("A BOX").
FT   DOMAIN       27     58       KRAB BOX ("B BOX").
FT   DOMAIN      108    298       ZINC-FINGERS I.
FT   DOMAIN      326    488       ZINC-FINGERS II.
FT   DOMAIN      503    721       ZINC-FINGERS III.
FT   DOMAIN      750    940       ZINC-FINGERS IV.
FT   DOMAIN      988   1066       ZINC-FINGERS V.
FT   DOMAIN     1136   1298       ZINC-FINGERS VI.
FT   ZN_FING     108    130       C2H2-TYPE.
FT   ZN_FING     136    158       C2H2-TYPE.
FT   ZN_FING     164    186       C2H2-TYPE.
FT   ZN_FING     192    214       C2H2-TYPE.
FT   ZN_FING     220    242       C2H2-TYPE.
FT   ZN_FING     248    270       C2H2-TYPE.
FT   ZN_FING     276    298       C2H2-TYPE.
FT   ZN_FING     326    348       C2H2-TYPE.
FT   ZN_FING     354    376       C2H2-TYPE.
FT   ZN_FING     382    404       C2H2-TYPE.
FT   ZN_FING     410    432       C2H2-TYPE.
FT   ZN_FING     438    460       C2H2-TYPE.
FT   ZN_FING     466    488       C2H2-TYPE.
FT   ZN_FING     503    525       C2H2-TYPE.
FT   ZN_FING     531    553       C2H2-TYPE.
FT   ZN_FING     559    581       C2H2-TYPE.
FT   ZN_FING     587    609       C2H2-TYPE.
FT   ZN_FING     615    637       C2H2-TYPE.
FT   ZN_FING     643    665       C2H2-TYPE.
FT   ZN_FING     671    693       C2H2-TYPE.
FT   ZN_FING     699    721       C2H2-TYPE.
FT   ZN_FING     750    772       C2H2-TYPE.
FT   ZN_FING     778    800       C2H2-TYPE.
FT   ZN_FING     806    828       C2H2-TYPE.
FT   ZN_FING     834    856       C2H2-TYPE.
FT   ZN_FING     862    884       C2H2-TYPE.
FT   ZN_FING     890    912       C2H2-TYPE.
FT   ZN_FING     918    940       C2H2-TYPE.
FT   ZN_FING     988   1010       C2H2-TYPE.
FT   ZN_FING    1016   1038       C2H2-TYPE.
FT   ZN_FING    1044   1066       C2H2-TYPE.
FT   ZN_FING    1136   1158       C2H2-TYPE.
FT   ZN_FING    1164   1186       C2H2-TYPE.
FT   ZN_FING    1192   1214       C2H2-TYPE.
FT   ZN_FING    1220   1242       C2H2-TYPE.
FT   ZN_FING    1248   1270       C2H2-TYPE.
FT   ZN_FING    1276   1298       C2H2-TYPE.
FT   STRAND     1045   1052
FT   STRAND     1052   1052
FT   HELIX      1056   1065
FT   TURN       1066   1066
SQ   SEQUENCE   1350 AA;  155804 MW;  1OCA7C8C CRC32;

Query Match             7.3%;  Score 109;  DB 1;  Length 1350;  .
Best Local Similarity  37.5%;  Pred. No. 4.14e-03;
Matches   12;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Db       1042  RPYKCGLCERSFVEKSALSRHQRVHKNESPVL 1073
               |:|:::|:|   ||  :: | |:  | ::::|
Qy         64  RSYTCSFCKREFRSAQALGGHMNVHRRDRAIL 95

RESULT    9
ID   ZN75_HUMAN     STANDARD;      PRT;      289 AA.
AC   P51815;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN 75.
GN   ZNF75.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG FIBROBLAST;
RX   MEDLINE; 94116987.
RA   VILLA A., ZUCCHI I., PILIA G., STRINA D., SUSANI L., MORALI F.,
RA   PATROSSO C., FRATTINI A., LUCCHINI F., REPETTO M., SACCO M.G.,
RA   ZOPPE M., VEZZONI P.;
RT   "ZNF75: isolation of a cDNA clone of the KRAB zinc finger gene
RT   subfamily mapped in YACs 1 Mb telomeric of HPRT.";
RL   GENOMICS 18:223-229(1993).
RN   [2]
RP   SEQUENCE OF 139-289 FROM N.A.
RA   MARINO M., ARCHIDIACONO N., FRANZE N., ROSATI M., ROCCHI M.,
RA   BALLABIO A., GRIMALDI G.;
RL   SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL SUBFAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 KRAB BOX.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----------------------------------------------------------------------
DR    EMBL; S67970; G460903; -.
DR    EMBL; X68010; G525244; -.
DR    MIM; 314997; -.
DR    PROSITE; PS00028; ZINC_FINGER_C2H2; 5.
DR    PFAM; PF00096; zf-C2H2; 5.
DR    HSSP; P25490; 1ZNM.
KW    TRANSCRIPTION REGULATION; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
FT    DOMAIN       16     55       KRAB BOX.
FT    DOMAIN      144    278       ZINC-FINGERS.
FT    ZN_FING     144    166       C2H2-TYPE.
FT    ZN_FING     172    194       C2H2-TYPE.
FT    ZN_FING     200    222       C2H2-TYPE.
FT    ZN_FING     228    250       C2H2-TYPE.
FT    ZN_FING     256    278       C2H2-TYPE.
SQ    SEQUENCE   289 AA;   33683 MW;  669284FB CRC32;

Query Match             7.2%;  Score 107;  DB 1;  Length 289;
Best Local Similarity  46.9%;  Pred. No. 8.39e-03;
Matches   15;  Conservative    7;  Mismatches  10;  Indels   0;  Gaps    0;

Db     254  QPYTCSLCKRNFSRRSSLLRHQKLHRRREACL 285
            :||||||||:|||| : |  ::||| ||
Qy      64  RSYTCSFCKREFRSAQALGGHMNVHRRDRAIL 95


RESULT   10
ID    BRC3_DROME    STANDARD;      PRT;     704 AA.
AC    Q01293;
DT    01-NOV-1995 (REL. 32, CREATED)
DT    01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT    01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE    BROAD-COMPLEX CORE-NS-Z3 PROTEIN (BRCORE-NS-23).
GN    BR OR BR-C.
OS    DROSOPHILA MELANOGASTER (FRUIT FLY).
OC    EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC    PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC    DROSOPHILIDAE; DROSOPHILA.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=OREGON-R;
RX    MEDLINE; 92077389.
RA    DIBELLO P.R., WITHERS D.A., BAYER C.A., FRISTROM J.W., GUILD G.M.;
RT    "The Drosophila Broad-Complex encodes a family of related proteins
RT    containing zinc fingers.";
RL    GENETICS 129:385-397(1991).
CC    -!- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND
CC        TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.
CC    -!- SUBCELLULAR LOCATION: NUCLEAR.
CC    -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR PRODUCTS ARE PRODUCED BY
CC        ALTERNATIVE SPLICING, BCORE-NS-23 (SHOWN HERE), BCORE-TNT1-Q1-Z1
CC        (AC Q01295), BCORE-Q1-Z1 (AC Q01295) AND BCORE-Z2 (AC Q01296). A
CC        COMMON CORE N-TERMINAL DOMAIN IS FUSED TO ALTERNATE ZINC-FINGER C-
CC        TERMINAL DOMAINS.
CC    -!- SIMILARITY: CONTAINS A N-TERMINAL BR-C/TTK DOMAIN.
CC    -----------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----------------------------------------------------------------------
DR    EMBL; X54664; G10901; -.
DR    FLYBASE; FBgn0000210; br.
DR    PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
DR    PFAM; PF00096; zf-C2H2; 2.
DR    PFAM; PF00651; BTB; 1.
DR    TRANSFAC; T01479; -.
KW    NUCLEAR PROTEIN; DNA-BINDING; DEVELOPMENTAL PROTEIN; ZINC-FINGER;
KW    METAL-BINDING; ALTERNATIVE SPLICING.
FT    DOMAIN        1    115       BR-C/TTK DOMAIN.
FT    DOMAIN      203    207       POLY-ALA.
FT    DOMAIN      265    268       POLY-ASN.
FT    DOMAIN      435    551       ASP/SER-RICH.
FT    DOMAIN      540    544       POLY-ASN.
FT    DOMAIN      545    551       POLY-SER.
FT    DOMAIN      587    640       ZINC-FINGERS.
FT    ZN_FING     587    610       C2H2-TYPE.
FT    ZN_FING     617    640       C2H2-TYPE.
SQ    SEQUENCE   704 AA;   74468 MW;  9CC18352 CRC32;

Query Match             7.2%;  Score 108;  DB 1;  Length 704;
Best Local Similarity  30.3%;  Pred. No. 5.90e-03;
Matches   10;  Conservative   12;  Mismatches  10;  Indels   10;  Gaps    1;

Db     617  YVCEFCHRRYRTKNSLTTHKSLQHRGSSGMLKR 649
            || ||  |:  || :::  :|::::|:   :||:
Qy      66  YTCSFCKREFRSAQALGGHMNVHRRDRA-ILRQ 97


RESULT   11
ID    SUHW_DROVI    STANDARD;      PRT;     899 AA.
AC    Q08876;
DT    01-OCT-1994 (REL. 30, CREATED)
DT    01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT    01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE    SUPPRESSOR OF HAIRY WING PROTEIN.
GN    SU(HW).
OS    DROSOPHILA VIRILIS (FRUIT FLY).
OC    EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC    PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC    DROSOPHILIDAE; DROSOPHILA.
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 94010293.
RA    HARRISON D.A., GDULA D.A., COYNE R.S., CORCES V.G.;
RT    "A leucine zipper domain of the suppressor of Hairy-wing protein
RT    mediates its repressive effect on enhancer function.";
RL    GENES DEV. 7:1966-1978(1993).
CC    -!- FUNCTION: SU(HW) CONTROLS THE PHENOTYPIC EFFECT OF THE GYPSY
CC        TRANSPOSABLE ELEMENT. BINDS SPECIFICALLY TO A REGION OF THE GYPSY
CC        ELEMENT LOCATED 3' OF THE 5'LTR. IT IS PROBABLY A TRANSCRIPTION
CC        FACTOR. COULD PLAY A ROLE IN THE ESTABLISHMENT OF CHROMATIN
CC        DOMAINS.
CC    -!- SUBCELLULAR LOCATION: NUCLEAR.
CC    -----------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----------------------------------------------------------------------
DR    EMBL; Z25520; G396677; -.
DR    PIR; S36436; S36436.
DR    PIR; B48586; B48586.
DR    FLYBASE; FBgn0013144; Dvir\su(Hw).
DR    PROSITE; PS00028; ZINC_FINGER_C2H2; 10.
DR    PFAM; PF00096; zf-C2H2; 12.
DR    HSSP; P25490; 1UBD.
KW    TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
KW    REPEAT; NUCLEAR PROTEIN.
FT    DOMAIN      218    617       ZINC-FINGERS.
FT    ZN_FING     218    240       C2H2-TYPE.
FT    ZN_FING     288    311       C2H2-TYPE.
FT    ZN_FING     318    340       C2H2-TYPE.
FT    ZN_FING     347    365       C2H2-TYPE.
```

```
FT   ZN_FING        379    401          C2H2-TYPE.
FT   ZN_FING        412    434          C2H2-TYPE.
FT   ZN_FING        440    462          C2H2-TYPE.
FT   ZN_FING        468    490          C2H2-TYPE.
FT   ZN_FING        496    518          C2H2-TYPE.
FT   ZN_FING        522    544          C2H2-TYPE.
FT   ZN_FING        552    576          C2H2-TYPE.
FT   ZN_FING        594    617          C2H2-TYPE.
SQ   SEQUENCE     899 AA;   100592 MW;   A884EA7A CRC32;

Query Match              7.2%;    Score 108;   DB 1;   Length 899;
Best Local Similarity   29.6%;    Pred. No. 5.90e-03;
Matches   16;   Conservative   16;   Mismatches 20;   Indels   2;   Gaps   2;

Db   520 RPYACSLCIQFKTEKQLERHVKDHTRQKRASFACTECTRSFRTSALLKEHLDA 573
           |:|:||| | | |  | ||::||  |  |   :|  : :|| |||:|
Qy    64 RSYTCSFCKREFRSAQALGGHMNVH-RRDRAILR-QSPPRDINRYSLLNLNLEP 115


RESULT  12
ID   ZG8_XENLA      STANDARD;      PRT;    196 AA.
AC   P18737;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE   GASTRULA ZINC FINGER PROTEIN XLCGF8.2DB (FRAGMENT).
OS   XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC   MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90040698.
RA   NIETFELD W., EL-BARADI T., MENTZEL H., PIELER T., KOESTER M.,
RA   POETING A., KNOECHEL W.;
RT   "Second-order repeats in Xenopus laevis finger proteins.";
RL   J. MOL. BIOL. 208:639-659(1989).
DR   PIR; S06561; S06561.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 7.
DR   PFAM; PF00096; zf-C2H2; 7.
DR   HSSP; P25490; 1ZNM.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT.
FT   NON_TER          1      1
FT   ZN_FING          6     28          C2H2-TYPE.
FT   ZN_FING         34     56          C2H2-TYPE.
FT   ZN_FING         62     84          C2H2-TYPE.
FT   ZN_FING         90    112          C2H2-TYPE.
FT   ZN_FING        118    140          C2H2-TYPE.
FT   ZN_FING        146    168          C2H2-TYPE.
FT   ZN_FING        174    196          C2H2-TYPE.
FT   NON_TER        196    196
SQ   SEQUENCE     196 AA;   22023 MW;   3799487C CRC32;

Query Match              7.0%;    Score 105;   DB 1;   Length 196;
Best Local Similarity   24.3%;    Pred. No. 1.69e-02;
Matches   17;   Conservative   25;   Mismatches 25;   Indels   3;   Gaps   3;

Db    98 EFAHKHRLLGGHLKIHTGEKPFSCTECGKHFAHKYHLVSHMKIHTREKPFTCTECGEHFAN 157
           :|:::  |||| |     ::|| |   :|||:| |:|: | |:|:|  : | : |::
Qy    47 NYGNEDHLLGGL-FSWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAIL-RQSPPRDIN 104

Db   158 KVSLLGHLKM 167
           :|||  |:|:
Qy   105 RYSLL-NLNL 113


RESULT  13
ID   SNAI_MOUSE     STANDARD;      PRT;    264 AA.
AC   Q02085;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   SNAIL PROTEIN HOMOLOG (SNA PROTEIN).
GN   SNA.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 93201990.
RA   SMITH D.E., DEL AMO F.F., GRIDLEY T.;
RT   "Isolation of Sna, a mouse gene homologous to the Drosophila genes
RT   snail and escargot: its expression pattern suggests multiple roles
RT   during postimplantation development.";
RL   DEVELOPMENT 116:1033-1039(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93130772.
RA   NIETO A.M., BENNETT M.F., SARGENT M.G., WILKINSON D.G.;
RT   "Cloning and developmental expression of Sna, a murine homologue of
RT   the Drosophila snail gene.";
RL   DEVELOPMENT 116:227-237(1992).
RN   [3]
RP   SEQUENCE OF 1-27 FROM N.A.
RC   STRAIN=129/SV;
RA   JIANG R., COPELAND N.G., GILBERT D.J., JENKINS N.A., GRIDLEY T.;
RL   SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: THIS PROTEIN HAS MANY ROLES DURING POSTIMPLANTATION
CC       DEVELOPMENT. IT IS INVOLVED IN EMBRYONIC MESODERM FORMATION
CC       AND ITS MAINTENANCE AND MAY ALSO BE INVOLVED IN CHONDROGENESIS
CC       AND IN EPITHELIAL-MESENCHYMAL INDUCTIVE INTERACTIONS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- DEVELOPMENTAL STAGE: POSTIMPLANTATION.
CC   -!- SIMILARITY: TO D.MELANOGASTER SNAIL PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M95604; G429188; -.
DR   EMBL; X67253; G54121; -.
DR   EMBL; U95961; G2105425; -.
DR   MGD; MGI:98330; SNA.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 3.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   HSSP; P08047; 1SP2.
DR   TRANSFAC; T02333; -.
KW   DEVELOPMENTAL PROTEIN; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
KW   NUCLEAR PROTEIN; REPEAT.
FT   DOMAIN         236    259          ZINC-FINGERS.
FT   ZN_FING        156    176          C2H2-TYPE.
FT   ZN_FING        180    202          C2H2-TYPE.
FT   ZN_FING        210    230          C2H2-TYPE.
FT   ZN_FING        236    259          C2H2-TYPE.
FT   CONFLICT       219    219          D -> V (IN REF. 2).
SQ   SEQUENCE     264 AA;   29190 MW;   692E6770 CRC32;

Query Match              7.0%;    Score 104;   DB 1;   Length 264;
Best Local Similarity   28.8%;    Pred. No. 2.38e-02;
Matches   15;   Conservative   15;   Mismatches 22;   Indels   0;   Gaps   0;

Db   206 KPFSCSCHCNRAFADRSNLRAHLQTHSDVKRYCQACARTFSRMSLLHKHQES 257
           :::| || :| | |:  |: :|||||:  :::   |  ||::| | :|:|:
Qy    64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINRYSLLNLNLEP 115


RESULT  14
ID   ODD_DROME      STANDARD;      PRT;    392 AA.
AC   P23803;
DT   01-NOV-1991 (REL. 20, CREATED)
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
```

```
DE   ODD-SKIPPED PROTEIN.
GN   ODD.
OS   DROSOPHILA MELANOGASTER (FRUIT FLY).
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC   DROSOPHILIDAE; DROSOPHILA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91006082.
RA   COULTER D.E., SWAYKUS E.A., BERAN-KOEHN M.A., GOLDBERG D.,
RA   WIESCHAUS E., SCHEDL P.;
RT   "Molecular analysis of odd-skipped, a zinc finger encoding
RT   segmentation gene with a novel pair-rule expression pattern.";
RL   EMBO J. 9:3795-3804(1990).
CC   -!- FUNCTION: PAIR-RULE PROTEIN THAT ESTABLISH PORTION OF ALTERNATING
CC       SEGMENTS DURING EMBRYOGENESIS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X57480; G296793; -.
DR   PIR; S11998; S11998.
DR   FLYBASE; FBgn0002985; odd.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
DR   PFAM; PF00096; zf-C2H2; 4.
DR   HSSP; P08153; 1ZFD.
DR   TRANSFAC; T00667; -.
KW   DEVELOPMENTAL PROTEIN; PAIR-RULE PROTEIN; ZINC-FINGER; METAL-BINDING;
KW   DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT   DOMAIN     82     93     POLY-GLN.
FT   DOMAIN    101    106     POLY-GLN.
FT   DOMAIN    142    149     POLY-GLN.
FT   DOMAIN    150    181     HIS/PRO-RICH.
FT   DOMAIN    220    326     ZINC-FINGERS.
FT   ZN_FING   220    242     C2H2-TYPE.
FT   ZN_FING   248    270     C2H2-TYPE.
FT   ZN_FING   276    298     C2H2-TYPE.
FT   ZN_FING   304    326     C2H2-TYPE.
FT   DOMAIN    369    375     POLY-SER.
SQ   SEQUENCE   392 AA;  44654 MW;  153B581E CRC32;

Query Match
Best Local Similarity   32.3%;   Score 105;   DB 1;   Length 392;
Matches   10;   Conservative   10;   Mismatches   11;   Indels   0;   Gaps   0;

Db    246 RPYSCDICGKAFRRQDHLRDHRYIHSKDKPF 276
Qy     64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAI 94


RESULT 15
ID   ZNG1_RAT     STANDARD;     PRT;     423 AA.
AC   Q07120;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN GFI-1 (GROWTH FACTOR INDEPENDENCE-1).
GN   GFI1 OR GFI-1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93180823.
RA   GILKS C.B., BEAR S.E., GRIMES H.L., TSICHLIS P.N.;
RT   "Progression of interleukin-2 (IL-2)-dependent rat T cell lymphoma
RT   lines to IL-2-independent growth following activation of a gene
RT   (Gfi-1) encoding a novel zinc finger protein.";
RL   MOL. CELL. BIOL. 13:1759-1768(1993).
CC   -!- FUNCTION: MAY BE A TRANSCRIPTION FACTOR INVOLVED IN REGULATING THE
CC       EXPRESSION OF GENES ACTIVE IN THE S PHASE DURING CELL CYCLE
CC       PROGRESSION IN T CELLS. MAY BE INVOLVED IN TUMOR PROGRESSION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO LYMPHOID TISSUES AND TESTES IN
CC       ADULT ANIMALS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION ENHANCED LATE AFTER INTERACTION OF
CC       IL-2 WITH ITS RECEPTOR. APPROXIMATELY WHEN CELLS ENTER S PHASE.
CC   -!- DOMAIN: THE GLY/ALA-RICH REGION MAY INDICATE THAT GFI-1 IS A
CC       REPRESSOR OF ITS TARGET GENES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L06986; G294553; -.
DR   PIR; A48152; A48152.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 6.
DR   PFAM; PF00096; zf-C2H2; 6.
DR   HSSP; P25490; 1ZNM.
DR   TRANSFAC; T01921; -.
KW   TRANSCRIPTION REGULATION; ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
FT   DOMAIN    159    210     GLY/ALA-RICH.
FT   DOMAIN    256    420     ZINC-FINGERS.
FT   ZN_FING   256    279     C2H2-TYPE.
FT   ZN_FING   285    307     C2H2-TYPE.
FT   ZN_FING   313    335     C2H2-TYPE.
FT   ZN_FING   341    363     C2H2-TYPE.
FT   ZN_FING   369    391     C2H2-TYPE.
FT   ZN_FING   397    420     C2H2-TYPE.
SQ   SEQUENCE   423 AA;  45857 MW;  5868AF54 CRC32;

Query Match
Best Local Similarity   24.4%;   Score 104;   DB 1;   Length 423;   Pred. No. 2.38e-02;
Matches   31;   Conservative   19;   Mismatches   23;   Indels   5;   Gaps   4;

Db    283 RPFACEMCGKTFGHAVSLEQHKAVHSQERSFDCKICGKSFKRSSTLSTHLLIHSDTRPYP 342
Qy     64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINRYSLLNLNL--EPNPNFYP 121

Db    343 CQY-C-GKRFHQKSDMKK 358
Qy    122 SHNPSFSRKFPP-FEMRK 138


Search completed: Wed Jul 14 17:25:10 1999
Job time : 10 secs.
```

```
  *********************************************
  *                                           *
  *            M P S R C H  (TM)              *
  *                                           *
  *********************************************

  Release 3.1A John F. Collins, Biocomputing Research Unit.
  Copyright (c) 1993-1998 University of Edinburgh, U.K.
  Distribution rights by Oxford Molecular Ltd
```

MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Jul 14 17:25:27 1999;   MasPar time 16.92 Seconds
                                               664.411 Million cell updates/sec

Tabular output not generated.

Title: >US-09-156-580-2
Description: (1-206) from US09156580A.pep
Perfect Score: 1495
Sequence: 1 MAGMDRNSFNSKYFKNKSIM..........IGLISESKEDLDLELRLGST 206

Scoring table: PAM 150
                  Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database: sptrembl9
                  1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
                  5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
                  9:sp_phage 10:sp_plant 12:sp_rodent 12:sp_unclassified
                  13:sp_vertebrate 14:sp_virus

Statistics: Mean 44.353; Variance 77.490; scale 0.572

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	364	24.3	204	10	Q38895	SUPERMAN.	1.10e-53
2	276	18.5	304	10	O80942	PUTATIVE ZINC FINGER P	1.01e-34
3	263	17.6	180	10	O23621	ZN FINGER PROTEIN HOMO	5.34e-32
4	169	11.3	209	10	Q39266	ZINC FINGER PROTEIN.	2.87e-13
5	164	11.0	228	10	Q42485	ZINC FINGER PROTEIN 1.	2.44e-12
6	160	10.7	197	10	Q39265	ZINC FINGER PROTEIN.	1.33e-11
7	157	10.5	259	10	Q39263	ZINC FINGER PROTEIN.	4.69e-11
8	154	10.3	150	10	Q39261	ZINC FINGER PROTEIN.	1.64e-10
9	148	9.9	235	10	Q39262	ZINC FINGER PROTEIN.	1.96e-09
10	141	9.4	239	10	Q39267	ZINC FINGER PROTEIN (F	3.37e-08
11	141	9.4	257	10	P93751	ZINC FINGER PROTEIN 8,	3.37e-08
12	132	8.8	166	10	O22086	ZPT2-14.	1.20e-06
13	129	8.6	341	10	P93714	PETHY, ZPT2-6.	3.86e-06
14	129	8.6	554	10	P93718	PETHY, ZPT4-2.	3.86e-06
15	128	8.6	554	10	P93719	TRANSCRIPTION FACTOR Z	5.67e-06
16	126	8.4	146	10	O22087	ZPT2-7.	1.22e-05
17	125	8.4	211	10	Q39264	ZINC FINGER PROTEIN.	1.79e-05
18	126	8.4	277	10	O40896	ZINC-FINGER DNA BINDIN	1.22e-05
19	126	8.4	368	5	Q21158	SIMILAR TO S. CEREVISI	1.22e-05
20	126	8.4	1186	11	O08961	ZINC FINGER PROTEIN 10	1.22e-05
21	123	8.2	274	10	Q40897	ZINC-FINGER DNA BINDIN	3.83e-05
22	121	8.1	238	10	O22533	PUTATIVE C2H2 ZINC FIN	8.13e-05
23	119	8.0	168	10	Q42453	ZINC FINGER PROTEIN.	1.72e-04
24	119	8.0	176	10	P93713	PETHY, ZPT2-5,	1.72e-04
25	119	8.0	227	10	Q42423	ZINC FINGER PROTEIN.	1.72e-04
26	119	8.0	227	10	Q96289	SALT-TOLERANCE ZINC FI	1.72e-04
27	119	8.0	273	10	O65150	OSMOTIC STRESS-INDUCED	1.72e-04
28	120	8.0	284	10	Q42375	ZINC FINGER PROTEIN.	1.18e-04
29	119	7.9	477	10	Q22091	ZPT4-4.	1.72e-04
30	118	7.9	163	10	O22088	ZPT2-8.	2.49e-04
31	118	7.9	235	10	O04176	ZINC-FINGER PROTEIN-1.	2.49e-04
32	117	7.8	237	10	O04177	ZINC-FINGER PROTEIN BC	3.60e-04
33	117	7.8	240	10	P93166	SCOF-1.	3.60e-04
34	116	7.8	253	10	Q40898	ZINC-FINGER DNA BINDIN	5.20e-04
35	116	7.8	281	10	Q43614	DNA-BINDING PROTEIN.	5.20e-04
36	117	7.8	282	10	O22083	ZPT2-11.	3.60e-04
37	115	7.7	2391	5	O27732	CARBAMOYL-PHOSPHATE SY	7.49e-04
38	113	7.6	162	10	Q42410	ZINC FINGER PROTEIN.	1.55e-03
39	114	7.6	354	11	Q62509	ZINC FINGER PROTEIN (F	1.08e-03
40	114	7.6	455	11	Q62511	ZINC FINGER PROTEIN PZ	1.08e-03
41	113	7.6	594	5	P91805	ARYLPHORIN GENE-SPECIF	1.55e-03
42	114	7.6	814	11	O70162	MYELOID ZINC FINGER PR	1.08e-03
43	114	7.6	908	5	Q19389	SIMILAR TO S. MUTANS B	1.08e-03
44	113	7.6	977	5	O44322	REVERSE TRANSCRIPTASE	1.55e-03
45	112	7.5	253	10	O22082	ZPT2-10.	2.22e-03

ALIGNMENTS

```
RESULT   1
ID   Q38895        PRELIMINARY;       PRT;       204 AA.
AC   Q38895;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   SUPERMAN.
GN   SUP.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA;
RX   MEDLINE; 96069740.
RA   SAKAI H., MEDRANO L.J., MEYEROWITZ E.M.;
RT   "Role of SUPERMAN in maintaining Arabidopsis floral whorl
RT   boundaries.";
RL   NATURE 378:199-203(1995); -.
DR   EMBL; U38946; G1079669; -.
DR   TRANSFAC; T02275; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16154; ARath:1523; mn16154.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE 204 AA;   23084 MW;   C359F53D CRC32;

Query Match
Best Local Similarity  24.3%;       Score 364;  DB 10;  Length 204;
Matches 80;  Conservative 34;  Mismatches 38;  Indels 21;  Gaps 11;

Db   33 DHDYLLG--FSWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSTPSP 90
          : ||| |||||||||||||||||||||||||||||||||| ||   |||
Qy   50 NEDHLLGGLFSWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAILR--QSP------- 99

Db   91 PYPNPNYSYSTMANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTK 150
           :  |: ||||||  | :|: :|| :| |: ::  |::|:  :|  ::||:  ||
Qy  100 P-RDIN-RYSLLNLNLEPN--P-NFYPSHNPSFSRKFPPFEMRKLG-KGV-VPNNHLKSA 152

Db  151 KSSLLVEAGEATRFTSKD-ACKILRNDEIISLELEIGLINESEQDLDLELRLG 202
          : ||:    |:  : : : |: |:||||| | ::|||||||||||||||||
Qy  153 RGRFGVEKIDSFM-QEKECTTTVIKKSEFLRLDLGIGLISESKEDLDLELRLG 204
```

RESULT 2
ID   O80942   PRELIMINARY;   PRT;   304 AA.
AC   O80942;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   PUTATIVE ZINC FINGER PROTEIN.
GN   F13M22.24.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., KAUL S., LIN X., KETCHUM K.A., CROSBY M.L.,
RA   BRANDON R.C., SYKES S.M., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.";
RL   SUBMITTED (JUN-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AC004684; G3236256; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   304 AA;   34125 MW;   879B713E CRC32;

Query Match   18.5%; Score 276; DB 10; Length 304;
Best Local Similarity   86.8%;   Pred. No. 1.01e-34;
Matches   33; Conservative   4; Mismatches   4; Indels   1; Gaps   0;

Db   34 WPPRSYTCSFCRREFKSAQALGGHMNVHRRDRARLKQA 71
              |||||||||:|||::|||||||||||||||||||||:|| |:|:
Qy   61 WPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT 3
ID   O23621   PRELIMINARY;   PRT;   180 AA.
AC   O23621;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZN FINGER PROTEIN HOMOLOG.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RP   EU ARABIDOPSIS SEQUENCING PROJECT, ESSA;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RX   MEDLINE; 98121113.
RA   BEVAN M., BANCROFT I., BENT E., LOVE K., GOODMAN H., DEAN C.,
RA   BERGKAMP R., DIRKSE W., VAN STAVEREN M., STIEKEMA W., DROST L.,
RA   RIDLEY P., HUDSON S.A., PATEL K., MURPHY G., PIFFANELLI P., WEDLER H.,
RA   WEDLER E., WAMBUTT R., WEITZENEGGER T., POHL T.M., TERRYN N.,
RA   GIELEN J., VILLARROEL R., DE CLERCK R., VAN MONTAGU M., LECHARNY A.,
RA   AUBORG S., GY I., KREIS M., LAO N., KAVANAGH T., HEMPEL S., KOTTER P.,
RA   ENTIAN K.D., RIEGER M., SCHAEFFER M., FUNK B., MUELLER-AUER S.,
RA   SILVEY M., JAMES R., MONTFORT A., PONS A., PUIGDOMENECH P., DOUKA A.,
RA   VOUKELATOU E., MILIONI D., HATZOPOULOS P., PIRAVANDI E., OBERMAIER B.,
RA   HILBERT H., DUESTERHOFT A., MOORES T., JONES J.D.G., ENEVA T.,
RA   PALME K., BENES V., RECHMAN S., ANSORGE W., COOKE R., BERGER C.,
RA   DELSENY M., VOET M., VOLCKAERT G., MEWES H.W., KLOSTERMAN S.,
RA   SCHUELLER C., CHALWATZIS N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   NATURE 391:485-488(1998).
DR   EMBL; Z97344; E327088; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   180 AA;   19788 MW;   D0D6D1B1 CRC32;

Query Match   17.6%; Score 263; DB 10; Length 180;
Best Local Similarity   86.5%;   Pred. No. 5.34e-32;
Matches   32; Conservative   3; Mismatches   2; Indels   0; Gaps   0;

Db   34 TWPPRSYTCNFCRREFRSAQALGGHMNVHRRDRASSR 70
              :||||||||:|||:||||||||||||||||||||||| |
Qy   60 SWPPRSYTCSFCKREFRSAQALGGHMNVHRRDRAILR 96

RESULT 4
ID   Q39266   PRELIMINARY;   PRT;   209 AA.
AC   Q39266; O04755;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN.
GN   ZFP7.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA; TISSUE=ROOT;
RX   MEDLINE; 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   DEWAR K., BUEHLER E., FENG J., KIM C., LI Y., SHINN P., SUN H.,
RA   CONWAY A., CONWAY A., KURTZ D., OJI O., OSBORNE B., SHEN Y.K.,
RA   TORIUMI M., VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A.,
RA   THEOLOGIS A., ECKER J.R.;
RL   SUBMITTED (DEC-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; L39650; G790685; -.
DR   EMBL; AC000103; G2213609; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16158; ARath:1523;mn16158.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   209 AA;   23166 MW;   BBAF75D2 CRC32;

Query Match   11.3%; Score 169; DB 10; Length 209;
Best Local Similarity   47.2%;   Pred. No. 2.87e-13;
Matches   17; Conservative   13; Mismatches   6; Indels   0; Gaps   0;

Db   56 PRVFSCNYCRRKFYSSQALGGHQHONAHKRERTMAKRA 91
              |: :||:|:| ||::::| |:|::|::::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT 5
ID   Q42485   PRELIMINARY;   PRT;   228 AA.
AC   Q42485;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN 1.
GN   ZFP1.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   GOODMAN H.M., GALLANT P., KEIFER-HIGGINS S., RUBENFIELD M.,
RA   CHURCH G.M.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA; TISSUE=ROOT;
RX   MEDLINE; 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
DR   EMBL; U53501; G1297186; -.
DR   EMBL; L39644; G790673; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16159; ARath;1523;mn16159.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE 228 AA; 25468 MW; CB1D768B CRC32;

Query Match    11.0%;  Score 164; DB 10;  Length 228;
Best Local Similarity   47.2%;  Pred. No. 2.44e-12;
Matches  17;  Conservative  12;  Mismatches  7;  Indels  0;  Gaps  0;

Db   65 PRVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRG 100
         || ::|::|:|||||||||::||::| :::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT 6
ID   Q39265      PRELIMINARY;      PRT;   197 AA.
AC   Q39265;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN.
GN   ZFP6.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA; TISSUE=ROOT;
RX   MEDLINE; 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
DR   EMBL; L39649; G790683; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16157; ARath;1523;mn16157.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE 197 AA; 21414 MW; E4AB47FC CRC32;

Query Match    10.7%;  Score 160; DB 10;  Length 197;
Best Local Similarity   47.2%;  Pred. No. 1.33e-11;
Matches  17;  Conservative  11;  Mismatches  8;  Indels  0;  Gaps  0;

Db   38 SRKYECQYCCREFANSQALGGHQNAHKKERQLLKRA 73
         :| | | ||| :|:||||| | |:::| :||:::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT 7
ID   Q39263      PRELIMINARY;      PRT;   259 AA.
AC   Q39263;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN.
GN   ZFP4.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA; TISSUE=ROOT;
RX   MEDLINE; 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
DR   EMBL; L39647; G790679; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16159; ARath;1523;mn16159.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE 259 AA; 28644 MW; 71FEFA6D CRC32;

Query Match    10.5%;  Score 157; DB 10;  Length 259;
Best Local Similarity   45.7%;  Pred. No. 4.69e-11;
Matches  16;  Conservative  12;  Mismatches  7;  Indels  0;  Gaps  0;

Db   83 RVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRA 117
         : :|:|::| |:||||||||::||::|  :||:::
Qy   64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT 8
ID   Q39261      PRELIMINARY;      PRT;   150 AA.
AC   Q39261;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN.
GN   ZFP2.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA; TISSUE=SILIQUE;
RX   MEDLINE; 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
DR   EMBL; L39645; G790675; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16155; ARath;1523;mn16155.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE 150 AA; 16956 MW; 47E6CB5E CRC32;

Query Match    10.3%;  Score 154; DB 10;  Length 150;
Best Local Similarity   31.9%;  Pred. No. 1.64e-10;
Matches  23;  Conservative  23;  Mismatches  23;  Indels  4;  Gaps  3;

Db   49 PRVFSCNYCQRKFYSSQALGGHQNAHKLERTLAKKS--RELFRSSN-TVDSDQ-PYPFSG 104
         || :|:::| |::||||||:||:: ::||::|:|::
Qy   63 PRSYTCSFCKREFRSAQALGGHMNVHRRDRAILRQSPPRDINRYSLLNLNLEPNFYPS 122
Db   105 RFELYGRGYQGF 116
          :: :|      |
Qy   123 HNPSFSRKFPPF 134

RESULT 9
ID   Q39262      PRELIMINARY;      PRT;   235 AA.
AC   Q39262;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN.
GN   ZFP3.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
```

```
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE: 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
DR   EMBL; L39646; G790677; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
DR   MENDEL; 16156; ARAth:1523;mn16156.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   235 AA;  26351 MW;  BEEC55BF CRC32;

Query Match             9.9%;  Score 148;  DB 10;  Length 235;
Best Local Similarity   45.5%;  Pred. No. 1.96e-09;
Matches   15;  Conservative   12;  Mismatches   6;  Indels   0;  Gaps   0;

Db         61 FSCNYCQRTFYSSQALGGHQNAHKRERTLAKRG 93
                : :::::||| | :|:||||||| :::| ::::
Qy         66 YTCSFCKREFRSAQALGGHMNVHRRDRAILRQS 98

RESULT 10
ID   Q39267        PRELIMINARY;    PRT;    239 AA.
AC   Q39267;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN (FRAGMENT).
GN   ZFP8.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LANDSBERG ERECTA; TISSUE=LEAF;
RX   MEDLINE: 95322589.
RA   TAGUE B.W., GOODMAN H.M.;
RT   "Characterization of a family of Arabidopsis zinc finger protein
RT   cDNAs.";
RL   PLANT MOL. BIOL. 28:267-279(1995).
DR   EMBL; L39651; G790687; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
FT   NON_TER       1      1
SQ   SEQUENCE   239 AA;  27781 MW;  48976A9F CRC32;

Query Match             9.4%;  Score 141;  DB 10;  Length 239;.
Best Local Similarity   51.7%;  Pred. No. 3.37e-08;
Matches   15;  Conservative   7;  Mismatches   7;  Indels   0;  Gaps   0;

Db         91 RRFECHYCFRNFPTSQALGGHQNAHKRER 119
                || : :||| :::||||||||| | :|:|
Qy         64 RSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 11
ID   P93751        PRELIMINARY;    PRT;    257 AA.
AC   P93751;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN 8, ZFP8.
GN   T06D20.16.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., PHILLIPS C.A., BRANDON R.C.,
RA   FUHRMANN J.L., WHITE O., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA   VENTER J.C.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; U90439; G1871188; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 1.
DR   PFAM; PF00096; zf-C2H2; 1.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   257 AA;  29571 MW;  FB769546 CRC32;

Query Match             9.4%;  Score 141;  DB 10;  Length 257;
Best Local Similarity   51.7%;  Pred. No. 3.37e-08;
Matches   15;  Conservative   7;  Mismatches   7;  Indels   0;  Gaps   0;

Db         93 RRFECHYCFRNFPTSQALGGHQNAHKRER 121
                || : :||| :::||||||||| | :|:|
Qy         64 RSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 12
ID   O22086        PRELIMINARY;    PRT;    166 AA.
AC   O22086;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   ZPT2-14.
OS   PETUNIA X HYBRIDA.
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC   ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; PETUNIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MITCHELL DIPLOID; TISSUE=ROOT;
RX   MEDLINE: 98083196.
RA   KUBO K., SAKAMOTO A., KOBAYASHI A., RYBKA Z., KANNO Y., NAKAGAWA H.,
RA   NISHINO T., TAKATSUJI H.;
RT   "Cys2/His2 zinc-finger protein family of petunia: evolution and
RT   general mechanism of target-sequence recognition.";
RL   NUCLEIC ACIDS RES. 26:608-615(1998).
DR   EMBL; AB006601; D1022782; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
DR   PFAM; PF00096; zf-C2H2; 2.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ   SEQUENCE   166 AA;  18922 MW;  2EF440AC CRC32;

Query Match             8.8%;  Score 132;  DB 10;  Length 166;
Best Local Similarity   32.6%;  Pred. No. 1.20e-06;
Matches   15;  Conservative   15;  Mismatches   15;  Indels   1;  Gaps   1;

Db         27 FSHQENHFNTMDNSPSRVFECKTCNRQFSSFQALGGHRASHKKPR 72
                ::::: :: :|:: | :::| | : ::| | |||||  :|:: :|
Qy         48 YGNEDHLLGGLF-SWPPRSYTCSFCKREFRSAQALGGHMNVHRRDR 92

RESULT 13
ID   P93714        PRELIMINARY;    PRT;    341 AA.
AC   P93714;
DT   01-MAY-1997 (TREMBLREL. 03, CREATED)
DT   01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   PETHY.ZPT2-6.
OS   PETUNIA X HYBRIDA.
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC   ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; PETUNIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=STAMEN; STRAIN=CV. MITCHELL DIPLOID;
RA   KOBAYASHI A., SAKAMOTO A., KUBO K., RYBKA Z., KANNO Y., TAKATSUJI H.;
RT   "Seven zinc-finger transcription factors are expressed sequentially
```

```
RT     during the development of anthers and pollen in Petunia.";
RL     SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR     EMBL; AB000452; D1019855; -.
DR     PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
DR     PFAM; PF00096; zf-C2H2; 2.
DR     MENDEL; 12063; PETx:1523:2.
KW     ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ     SEQUENCE   341 AA;  38431 MW;  F10F3662 CRC32;

Query Match             8.6%;  Score 129;  DB 10;  Length 341;
Best Local Similarity  60.7%;  Pred. No. 3.86e-06;
Matches   17;  Conservative    3;  Mismatches    8;  Indels    0;  Gaps    0;

Db   236 SPRIHECSYCGAEFTSGQALGGHMRRHR 263
             :|  |  | | |:|||||||| ||
Qy    62 PPRSYTCSFCKREFRSAQALGGHMNVHR 89


RESULT   14
ID     P93718           PRELIMINARY;      PRT;    554 AA.
AC     P93718;
DT     01-MAY-1997 (TREMBLREL. 03, CREATED)
DT     01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT     01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE     PETHY, ZPT4-2.
OS     PETUNIA X HYBRIDA.
OC     EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC     EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC     ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; PETUNIA.
RN     [1]
RP     SEQUENCE FROM N.A.
RC     TISSUE=STAMEN; STRAIN=CV. MITCHELL DIPLOID;
RA     KOBAYASHI A., SAKAMOTO A., KUBO K., RYBKA Z., KANNO Y., TAKATSUJI H.;
RT     "Seven zinc-finger transcription factors are expressed sequentially
RT     during the development of anthers and pollen in Petunia.";
RL     SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR     EMBL; AB000456; D1020715; -.
DR     PROSITE; PS00028; ZINC_FINGER_C2H2; 3.
DR     PFAM; PF00096; zf-C2H2; 3.
DR     MENDEL; 12067; PETx:1523:6.
KW     ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ     SEQUENCE   554 AA;  62932 MW;  8C0AA70D CRC32;

Query Match             8.6%;  Score 129;  DB 10;  Length 554;
Best Local Similarity  38.2%;  Pred. No. 3.86e-06;
Matches   21;  Conservative   15;  Mismatches   17;  Indels    2;  Gaps    2;

Db   488 KGHACFPCPRMFKSGQALGGHKRSHFIVSSENHYQASAVQGKIVDLLDLNL-PAP 541
             :: :|| :|:::|||||| |   ::  : :  | |  :    |:||| ||
Qy    64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAILR-QSPPRDINRYSLLNLNLEPNP 117


RESULT   15
ID     P93719           PRELIMINARY;      PRT;    554 AA.
AC     P93719; O04749;
DT     01-MAY-1997 (TREMBLREL. 03, CREATED)
DT     01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT     01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE     TRANSCRIPTION FACTOR ZPT4-3.
OS     PETUNIA X HYBRIDA (GARDEN PETUNIA).
OC     EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC     EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC     ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; PETUNIA.
RN     [1]
RP     SEQUENCE FROM N.A.
RC     STRAIN=CV. MITCHELL DIPLOID; TISSUE=STAMEN;
RA     KOBAYASHI A., SAKAMOTO A., KUBO K., RYBKA Z., KANNO Y., TAKATSUJI H.;
RL     SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC     -!- SUBCELLULAR LOCATION: NUCLEAR.
DR     EMBL; AB003672; D1020953; -.
DR     PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
DR     PFAM; PF00096; zf-C2H2; 3.
DR     MENDEL; 12068; PETx:1523:7.
```

```
KW     TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
KW     NUCLEAR PROTEIN.
FT     ZN_FING         9        31        C2H2-TYPE.
FT     ZN_FING       119       141        C2H2-TYPE.
FT     ZN_FING       417       439        C2H2-TYPE.
FT     ZN_FING       488       510        C2H2-TYPE.
SQ     SEQUENCE   554 AA;  62616 MW;  E7072B50 CRC32;

Query Match             8.6%;  Score 128;  DB 10;  Length 554;
Best Local Similarity  39.3%;  Pred. No. 5.67e-06;
Matches   22;  Conservative   13;  Mismatches   18;  Indels    3;  Gaps    2;

Db   486 KGHECPYCDRVFKSGQALGGHKRSHFIGGSFRNLNQSSAAKKEADDLLDLNL-PAP 540
             :: |:|| | |:|:|||||||  |   :   ::  :  |!||  : |:||| ||
Qy    64 RSYTCSFCKREFRSAQALGGHMNVHRRDRAI--LRQSPPRDINRYSLLNLNLEPNP 117


Search completed: Wed Jul 14 17:25:46 1999
Job time : 19 secs.
```

THIS PAGE BLANK (USPTO)